

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

**UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office**

February 20, 2004

**THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM
THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK
OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT
APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A
FILING DATE.**

APPLICATION NUMBER: 60/484,584

FILING DATE: June 30, 2003

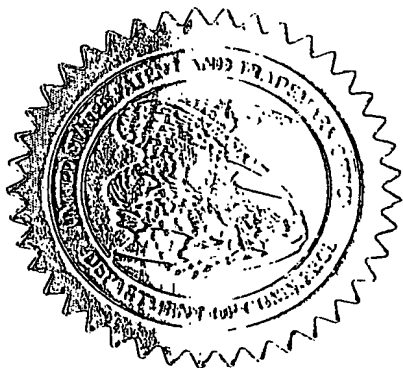
RELATED PCT APPLICATION NUMBER: PCT/US03/38855


RECEIVED

26 FEB 2004

WIPO PCT

**By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS**




P. SWAIN
Certifying Officer

**PRIORITY
DOCUMENT**

**SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1 (a) OR (b)**

BEST AVAILABLE COPY

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53 (c).

Express Mail Label No. ET 096 027 002 US

INVENTOR(S)					
Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)			
Shu-Hui Roberto	Liu Macina	Redwood City, CA San Jose, CA			
<input type="checkbox"/> Additional inventors are being named on the _____ separately numbered sheets attached hereto					
TITLE OF THE INVENTION (280 characters max)					
SPICE VARIANTS AND METHODS RELATING TO OVARIAN SPECIFIC GENES AND PROTEINS					
CORRESPONDENCE ADDRESS					
Direct all correspondence to:					
<input type="checkbox"/> Customer Number <input type="text"/> → <div>Place Customer Number Bar Code Label here</div>					
OR Type Customer Number here					
<input checked="" type="checkbox"/> Firm or Individual Name		Licata & Tyrrell			
Address		66 East Main Street			
Address					
City		Marlton	State	New Jersey	ZIP 08053
Country		USA	Telephone	(805) 810-1515	Fax (805) 810-1454
ENCLOSED APPLICATION PARTS (check all that apply)					
<input checked="" type="checkbox"/> Specification Total Number of Pages		162	<input type="checkbox"/> CD(s), Number		<input type="text"/>
<input checked="" type="checkbox"/> Drawing(s) Number of Sheets		29	<input type="checkbox"/> Other (specify)		<input type="text"/>
<input type="checkbox"/> Application Data Sheet. See 37 CFR 1.76					
METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT					
<input type="checkbox"/> Applicant claims small entity status. See 37 CFR 1.27.					
<input type="checkbox"/> A check or money order is enclosed to cover the filing fees					
FILING FEE AMOUNT (\$)					
<input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: 50-0434 160.00					
<input type="checkbox"/> Payment by credit card. Form PTO-2038 is attached.					
The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.					
<input checked="" type="checkbox"/> No.					
<input type="checkbox"/> Yes, the name of the U.S. Government agency and the Government contract number are: _____					

Respectfully submitted,

SIGNATURE

TYPED or PRINTED NAME

TELEPHONE

Nathan P. Letts

(650) 246-6400

Date:

REGISTRATION NO.

(if appropriate)

Docket Number:

30 June 2003

36,581

DEX-0439

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

This collection of information is required by 37 CFR 1.51. The information is used by the public to file (and by the PTO to process) a provisional application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 8 hours to complete, including gathering, preparing, and submitting the complete provisional application to the PTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, Washington, D.C., 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Box Provisional Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

16015 U.S. PTO

60/484584

DEX - 0439

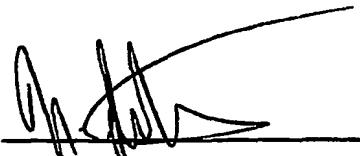
CERTIFICATE OF EXPRESS MAILING

"Express Mail" Label No. ET 096 027 002 US

Date of Deposit: 30 June 2003

I hereby certify that this paper is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R 1.10 on the date indicated above and is addressed to the "BOX PROVISIONAL", Assistant Commissioner for Patents, Washington, D.C. 20231.

- 1) Provision Application Transmittal Letter, with authorization to charge deposit account 50-0434 \$160.00 for filing fee (2 copies);**
- 2) Application consisting of 162 pages including; 156 pages of Specification including, 5 pages of Claims, 1 page of Abstract, with 29 pages of Figures; and**
- 3) Return Post Card.**



NATHAN P. LETTS
Registration No. 36,581

**SPLICE VARIANTS AND METHODS
RELATING TO OVARIAN SPECIFIC GENES AND PROTEINS**

FIELD OF THE INVENTION

5 The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic ovarian cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the
10 nucleic acids, polypeptides, antibodies, post translational modifications (PTMs), variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating ovarian cancer and non-cancerous disease states in ovarian, identifying ovarian tissue and monitoring and identifying and/or designing agonists and antagonists of
15 polypeptides of the invention. The uses also include gene therapy, therapeutic molecules including but limited to antibodies or antisense molecules, production of transgenic animals and cells, and production of engineered ovarian tissue for treatment and research.

BACKGROUND OF THE INVENTION

 Cancer of the ovaries is the fourth-most common cause of cancer death in women
20 in the United States, with more than 23,000 new cases and roughly 14,000 deaths predicted for the year 2001. Shridhar, V. et al., *Cancer Res.* 61(15): 5895-904 (2001); Memarzadeh, S. & Berek, J. S., *J. Reprod. Med.* 46(7): 621-29 (2001). The incidence of ovarian cancer is of serious concern worldwide, with an estimated 191,000 new cases predicted annually. Runnebaum, I. B. & Stickeler, E., *J. Cancer Res.*
25 *Clin. Oncol.* 127(2): 73-79 (2001). Unfortunately, women with ovarian cancer are typically asymptomatic until the disease has metastasized. Because effective screening for ovarian cancer is not available, roughly 70% of women diagnosed have an advanced stage of the cancer with a five-year survival rate of ~25-30%. Memarzadeh, S. & Berek, J. S., *supra*; Nunns, D. et al., *Obstet. Gynecol. Surv.* 55(12): 746-51. Conversely,
30 women diagnosed with early stage ovarian cancer enjoy considerably higher survival rates. Werness, B. A. & Eltabbakh, G. H., *Int'l. J. Gynecol. Pathol.* 20(1): 48-63 (2001). Although our understanding of the etiology of ovarian cancer is incomplete, the results of extensive research in this area point to a combination of age, genetics,

reproductive, and dietary/environmental factors. Age is a key risk factor in the development of ovarian cancer: while the risk for developing ovarian cancer before the age of 30 is slim, the incidence of ovarian cancer rises linearly between ages 30 to 50, increasing at a slower rate thereafter, with the highest incidence being among
5 septagenarian women. Jeanne M. Schilder et al., Hereditary Ovarian Cancer: Clinical Syndromes and Management, in Ovarian Cancer 182 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001).

With respect to genetic factors, a family history of ovarian cancer is the most significant risk factor in the development of the disease, with that risk depending on the
10 number of affected family members, the degree of their relationship to the woman, and which particular first degree relatives are affected by the disease. Id. Mutations in several genes have been associated with ovarian cancer, including BRCA1 and BRCA2, both of which play a key role in the development of breast cancer, as well as hMSH2 and hMLH1, both of which are associated with hereditary non-polyposis colon cancer. Katherine Y.
15 Look, Epidemiology, Etiology, and Screening of Ovarian Cancer, in Ovarian Cancer 169, 171-73 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001). BRCA1, located on chromosome 17, and BRCA2, located on chromosome 13, are tumor suppressor genes implicated in DNA repair; mutations in these genes are linked to roughly 10% of ovarian cancers. Id. at 171-72; Schilder et al., supra at 185-86. hMSH2 and hMLH1 are
20 associated with DNA mismatch repair, and are located on chromosomes 2 and 3, respectively; it has been reported that roughly 3% of hereditary ovarian carcinomas are due to mutations in these genes. Look, supra at 173; Schilder et al., supra at 184, 188-89.

Reproductive factors have also been associated with an increased or reduced risk of ovarian cancer. Late menopause, nulliparity, and early age at menarche have all been
25 linked with an elevated risk of ovarian cancer. Schilder et al., supra at 182. One theory hypothesizes that these factors increase the number of ovulatory cycles over the course of a woman's life, leading to "incessant ovulation," which is thought to be the primary cause of mutations to the ovarian epithelium. Id.; Laura J. Havrilesky & Andrew Berchuck, Molecular Alterations in Sporadic Ovarian Cancer, in Ovarian Cancer 25 (Stephen C.
30 Rubin & Gregory P. Sutton eds., 2d ed. 2001). The mutations may be explained by the fact that ovulation results in the destruction and repair of that epithelium, necessitating increased cell division, thereby increasing the possibility that an undetected mutation will occur. Id. Support for this theory may be found in the fact pregnancy, lactation, and the

use of oral contraceptives, all of which suppress ovulation, confer a protective effect with respect to developing ovarian cancer. *Id.*

Among dietary/environmental factors, there would appear to be an association between high intake of animal fat or red meat and ovarian cancer, while the antioxidant
5 Vitamin A, which prevents free radical formation and also assists in maintaining normal cellular differentiation, may offer a protective effect. Look, *supra* at 169. Reports have also associated asbestos and hydrous magnesium trisilicate (talc), the latter of which may be present in diaphragms and sanitary napkins. *Id.* at 169-70.

Current screening procedures for ovarian cancer, while of some utility, are quite
10 limited in their diagnostic ability, a problem that is particularly acute at early stages of cancer progression when the disease is typically asymptomatic yet is most readily treated. Walter J. Burdette, Cancer: Etiology, Diagnosis, and Treatment 166 (1998); Memarzadeh & Berek, *supra*; Runnebaum & Stickeler, *supra*; Werness & Eltabbakh, *supra*. Commonly used screening tests include biannual rectovaginal pelvic examination,
15 radioimmunoassay to detect the CA-125 serum tumor marker, and transvaginal ultrasonography. Burdette, *supra* at 166.

Pelvic examination has failed to yield adequate numbers of early diagnoses, and the other methods are not sufficiently accurate. *Id.* One study reported that only 15% of patients who suffered from ovarian cancer were diagnosed with the disease at the time of
20 their pelvic examination. Look, *supra* at 174. Moreover, the CA-125 test is prone to giving false positives in pre-menopausal women and has been reported to be of low predictive value in post-menopausal women. *Id.* at 174-75. Although transvaginal ultrasonography is now the preferred procedure for screening for ovarian cancer, it is unable to distinguish reliably between benign and malignant tumors, and also cannot
25 locate primary peritoneal malignancies or ovarian cancer if the ovary size is normal. Schilder et al., *supra* at 194-95. While genetic testing for mutations of the BRCA1, BRCA2, hMSH2, and hMLH1 genes is now available, these tests may be too costly for some patients and may also yield false negative or indeterminate results. Schilder et al., *supra* at 191-94.

30 The staging of ovarian cancer, which is accomplished through surgical exploration, is crucial in determining the course of treatment and management of the disease. AJCC Cancer Staging Handbook 187 (Irvin D. Fleming et al. eds., 5th ed. 1998); Burdette, *supra* at 170; Memarzadeh & Berek, *supra*; Shridhar et al., *supra*. Staging is performed

by reference to the classification system developed by the International Federation of Gynecology and Obstetrics. David H. Moore, Primary Surgical Management of Early Epithelial Ovarian Carcinoma, in Ovarian Cancer 203 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001); Fleming et al. eds., supra at 188. Stage I ovarian cancer is
5 characterized by tumor growth that is limited to the ovaries and is comprised of three substages. Id. In substage IA, tumor growth is limited to one ovary, there is no tumor on the external surface of the ovary, the ovarian capsule is intact, and no malignant cells are present in ascites or peritoneal washings. Id. Substage IB is identical to A1, except that tumor growth is limited to both ovaries. Id. Substage IC refers to the presence of tumor
10 growth limited to one or both ovaries, and also includes one or more of the following characteristics: capsule rupture, tumor growth on the surface of one or both ovaries, and malignant cells present in ascites or peritoneal washings. Id.

Stage II ovarian cancer refers to tumor growth involving one or both ovaries, along with pelvic extension. Id. Substage IIA involves extension and/or implants on the uterus
15 and/or fallopian tubes, with no malignant cells in the ascites or peritoneal washings, while substage IIB involves extension into other pelvic organs and tissues, again with no malignant cells in the ascites or peritoneal washings. Id. Substage IIC involves pelvic extension as in IIA or IIB, but with malignant cells in the ascites or peritoneal washings. Id.

20 Stage III ovarian cancer involves tumor growth in one or both ovaries, with peritoneal metastasis beyond the pelvis confirmed by microscope and/or metastasis in the regional lymph nodes. Id. Substage IIIA is characterized by microscopic peritoneal metastasis outside the pelvis, with substage IIIB involving macroscopic peritoneal metastasis outside the pelvis 2 cm or less in greatest dimension. Id. Substage IIIC is
25 identical to IIIB, except that the metastasis is greater than 2 cm in greatest dimension and may include regional lymph node metastasis. Id. Lastly, Stage IV refers to the presence distant metastasis, excluding peritoneal metastasis. Id.

While surgical staging is currently the benchmark for assessing the management and treatment of ovarian cancer, it suffers from considerable drawbacks, including the
30 invasiveness of the procedure, the potential for complications, as well as the potential for inaccuracy. Moore, supra at 206-208, 213. In view of these limitations, attention has turned to developing alternative staging methodologies through understanding differential gene expression in various stages of ovarian cancer and by obtaining various biomarkers

to help better assess the progression of the disease. Vartiainen, J. et al., *Int'l J. Cancer*, 95(5): 313-16 (2001); Shridhar et al. *supra*; Backelandt, M. et al., *J. Clin. Oncol.* 18(22): 3775-81.

The treatment of ovarian cancer typically involves a multiprong attack, with surgical intervention serving as the foundation of treatment. Dennis S. Chi & William J. Hoskins, *Primary Surgical Management of Advanced Epithelial Ovarian Cancer*, in Ovarian Cancer 241 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001). For example, in the case of epithelial ovarian cancer, which accounts for ~90% of cases of ovarian cancer, treatment typically consists of: (1) cytoreductive surgery, including total abdominal hysterectomy, bilateral salpingo-oophorectomy, omentectomy, and lymphadenectomy, followed by (2) adjuvant chemotherapy with paclitaxel and either cisplatin or carboplatin. Eltabbakh, G.H. & Awtrey, C.S., *Expert Op. Pharmacother.* 2(10): 109-24. Despite a clinical response rate of 80% to the adjuvant therapy, most patients experience tumor recurrence within three years of treatment. Id. Certain patients may undergo a second cytoreductive surgery and/or second-line chemotherapy. Memarzadeh & Berek, *supra*.

From the foregoing, it is clear that procedures used for detecting, diagnosing, monitoring, staging, prognosticating, and preventing the recurrence of ovarian cancer are of critical importance to the outcome of the patient. Moreover, current procedures, while helpful in each of these analyses, are limited by their specificity, sensitivity, invasiveness, and/or their cost. As such, highly specific and sensitive procedures that would operate by way of detecting novel markers in cells, tissues, or bodily fluids, with minimal invasiveness and at a reasonable cost, would be highly desirable.

Breast cancer, also referred to as mammary tumor cancer, is the second most common cancer among women, accounting for a third of the cancers diagnosed in the United States. One in nine women will develop breast cancer in her lifetime and about 192,000 new cases of breast cancer are diagnosed annually with about 42,000 deaths. Bevers, *Primary Prevention of Breast Cancer*, in Breast Cancer, 20-54 (Kelly K Hunt et al., ed., 2001); Kochanek et al., 49 *Nat'l. Vital Statistics Reports* 1, 14 (2001). Breast cancer is extremely rare in women younger than 20 and is very rare in women under 30. The incidence of breast cancer rises with age and becomes significant by age 50. White Non-Hispanic women have the highest incidence rate for breast cancer and Korean women have the lowest. Increased prevalence of the genetic mutations BRCA1 and BRCA2 that

promote breast and other cancers are found in Ashkenazi Jews. African American women have the highest mortality rate for breast cancer among these same groups (31 per 100,000), while Chinese women have the lowest at 11 per 100,000. Although men can get breast cancer, this is extremely rare. (American Cancer Society Website:

5 <http://www.cancer.org>). With the exception of those cases with associated genetic factors, precise causes of breast cancer are not known.

In the treatment of breast cancer, there is considerable emphasis on detection and risk assessment because early and accurate staging of breast cancer has a significant impact on survival. For example, breast cancer detected at an early stage (stage T0,
10 discussed below) has a five-year survival rate of 92%. Conversely, if the cancer is not detected until a late stage (i.e., stage T4 (IV)), the five-year survival rate is reduced to 13%. AJCC Cancer Staging Handbook pp. 164-65 (Irvin D. Fleming et al. eds., 5th ed. 1998). Some detection techniques, such as mammography and biopsy, involve increased discomfort, expense, and/or radiation, and are only prescribed only to patients with an
15 increased risk of breast cancer.

Current methods for predicting or detecting breast cancer risk are not optimal. One method for predicting the relative risk of breast cancer is by examining a patient's risk factors and pursuing aggressive diagnostic and treatment regimens for high risk patients. A patient's risk of breast cancer has been positively associated with increasing age,
20 nulliparity, family history of breast cancer, personal history of breast cancer, early menarche, late menopause, late age of first full term pregnancy, prior proliferative breast disease, irradiation of the breast at an early age and a personal history of malignancy. Lifestyle factors such as fat consumption, alcohol consumption, education, and socioeconomic status have also been associated with an increased incidence of breast
25 cancer although a direct cause and effect relationship has not been established. While these risk factors are statistically significant, their weak association with breast cancer limited their usefulness. Most women who develop breast cancer have none of the risk factors listed above, other than the risk that comes with growing older. NIH Publication No. 00-1556 (2000).

30 Current screening methods for detecting cancer, such as breast self exam, ultrasound, and mammography have drawbacks that reduce their effectiveness or prevent their widespread adoption. Breast self exams, while useful, are unreliable for the detection of breast cancer in the initial stages where the tumor is small and difficult to detect by

palpation. Ultrasound measurements require skilled operators at an increased expense. Mammography, while sensitive, is subject to over diagnosis in the detection of lesions that have questionable malignant potential. There is also the fear of the radiation used in mammography because prior chest radiation is a factor associated with an increase
5 incidence of breast cancer.

At this time, there are no adequate methods of breast cancer prevention. The current methods of breast cancer prevention involve prophylactic mastectomy (mastectomy performed before cancer diagnosis) and chemoprevention (chemotherapy before cancer diagnosis) which are drastic measures that limit their adoption even among
10 women with increased risk of breast cancer. Bevers, supra.

A number of genetic markers have been associated with breast cancer. Examples of these markers include carcinoembryonic antigen (CEA) (Mughal et al., JAMA 249:1881 (1983)), MUC-1 (Frische and Liu, J. Clin. Ligand 22:320 (2000)), HER-2/neu (Haris et al., Proc.Am.Soc.Clin.Oncology 15:A96 (1996)), uPA, PAI-1, LPA, LPC, RAK
15 and BRCA (Esteva and Fritsche, Serum and Tissue Markers for Breast Cancer, in Breast Cancer, 286-308 (2001)). These markers have problems with limited sensitivity, low correlation, and false negatives which limit their use for initial diagnosis. For example, while the BRCA1 gene mutation is useful as an indicator of an increased risk for breast cancer, it has limited use in cancer diagnosis because only 6.2 % of breast cancers are
20 BRCA1 positive. Malone et al., JAMA 279:922 (1998). See also, Mewman et al., JAMA 279:915 (1998) (correlation of only 3.3%).

There are four primary classifications of breast cancer varying by the site of origin and the extent of disease development.

- I. Ductal carcinoma in situ (DCIS): Malignant transformation of ductal
25 epithelial cells that remain in their normal position. DCIS is a purely localized disease, incapable of metastasis.
- II. Invasive ductal carcinoma (IDC): Malignancy of the ductal epithelial cells breaking through the basal membrane and into the supporting tissue of the breast. IDC may eventually spread elsewhere in the body.
- 30 III. Lobular carcinoma in situ (LCIS): Malignancy arising in a single lobule of the breast that fail to extend through the lobule wall, it generally remains localized.
- IV. Infiltrating lobular carcinoma (ILC): Malignancy arising in a single lobule of the breast and invading directly through the lobule wall into adjacent tissues.

By virtue of its invasion beyond the lobule wall, ILC may penetrate lymphatics and blood vessels and spread to distant sites.

For purpose of determining prognosis and treatment, these four breast cancer types have been staged according to the size of the primary tumor (T), the involvement of lymph nodes (N), and the presence of metastasis (M). Although DCIS by definition represents localized stage I disease, the other forms of breast cancer may range from stage II to stage IV. There are additional prognostic factors that further serve to guide surgical and medical intervention. The most common ones are total number of lymph nodes involved, ER (estrogen receptor) status, Her2/neu receptor status and histologic grades.

Breast cancers are diagnosed into the appropriate stage categories recognizing that different treatments are more effective for different stages of cancer. Stage TX indicates that primary tumor cannot be assessed (i.e., tumor was removed or breast tissue was removed). Stage T0 is characterized by abnormalities such as hyperplasia but with no evidence of primary tumor. Stage Tis is characterized by carcinoma in situ, intraductal carcinoma, lobular carcinoma in situ, or Paget's disease of the nipple with no tumor. Stage T1 (I) is characterized as having a tumor of 2 cm or less in the greatest dimension. Within stage T1, Tmic indicates microinvasion of 0.1 cm or less, T1a indicates a tumor of between 0.1 to 0.5 cm, T1b indicates a tumor of between 0.5 to 1 cm, and T1c indicates tumors of between 1 cm to 2 cm. Stage T2 (II) is characterized by tumors from 2 cm to 5 cm in the greatest dimension. Tumors greater than 5 cm in size are classified as stage T3 (III). Stage T4 (IV) indicates a tumor of any size with extension to the chest wall or skin. Within stage T4, T4a indicates extension of the tumor to the chest wall, T4b indicates edema or ulceration of the skin of the breast or satellite skin nodules confined to the same breast, T4c indicates a combination of T4a and T4b, and T4d indicates inflammatory carcinoma. AJCC Cancer Staging Handbook pp. 159-70 (Irvin D. Fleming et al. eds., 5th ed. 1998). In addition to standard staging, breast tumors may be classified according to their estrogen receptor and progesterone receptor protein status. Fisher et al., *Breast Cancer Research and Treatment* 7:147 (1986). Additional pathological status, such as HER2/neu status may also be useful. Thor et al., *J.Nat'l.Cancer Inst.* 90:1346 (1998); Paik et al., *J.Nat'l.Cancer Inst.* 90:1361 (1998); Hutchins et al., *Proc.Am.Soc.Clin.Oncology* 17:A2 (1998); and Simpson et al., *J.Clin.Oncology* 18:2059 (2000).

In addition to the staging of the primary tumor, breast cancer metastases to regional lymph nodes may be staged. Stage NX indicates that the lymph nodes cannot be assessed (e.g., previously removed). Stage N0 indicates no regional lymph node metastasis. Stage N1 indicates metastasis to movable ipsilateral axillary lymph nodes. 5 Stage N2 indicates metastasis to ipsilateral axillary lymph nodes fixed to one another or to other structures. Stage N3 indicates metastasis to ipsilateral internal mammary lymph nodes. Id.

Stage determination has potential prognostic value and provides criteria for designing optimal therapy. Simpson et al., J. Clin. Oncology 18:2059 (2000).

- 10 Generally, pathological staging of breast cancer is preferable to clinical staging because the former gives a more accurate prognosis. However, clinical staging would be preferred if it were as accurate as pathological staging because it does not depend on an invasive procedure to obtain tissue for pathological evaluation. Staging of breast cancer would be improved by detecting new markers in cells, tissues, or bodily fluids which could 15 differentiate between different stages of invasion. Progress in this field will allow more rapid and reliable method for treating breast cancer patients.

- Treatment of breast cancer is generally decided after an accurate staging of the primary tumor. Primary treatment options include breast conserving therapy (lumpectomy, breast irradiation, and surgical staging of the axilla), and modified radical 20 mastectomy. Additional treatments include chemotherapy, regional irradiation, and, in extreme cases, terminating estrogen production by ovarian ablation.

- Until recently, the customary treatment for all breast cancer was mastectomy. Fonseca et al., Annals of Internal Medicine 127:1013 (1997). However, recent data indicate that less radical procedures may be equally effective, in terms of survival, for 25 early stage breast cancer. Fisher et al., J. of Clinical Oncology 16:441 (1998). The treatment options for a patient with early stage breast cancer (i.e., stage Tis) may be breast-sparing surgery followed by localized radiation therapy at the breast. Alternatively, mastectomy optionally coupled with radiation or breast reconstruction may be employed. These treatment methods are equally effective in the early stages of breast cancer.

- 30 Patients with stage I and stage II breast cancer require surgery with chemotherapy and/or hormonal therapy. Surgery is of limited use in Stage III and stage IV patients. Thus, these patients are better candidates for chemotherapy and radiation therapy with surgery limited to biopsy to permit initial staging or subsequent restaging because cancer

is rarely curative at this stage of the disease. AJCC Cancer Staging Handbook 84, 164-65 (Irvin D. Fleming *et al.* eds., 5th ed.1998).

In an effort to provide more treatment options to patients, efforts are underway to define an earlier stage of breast cancer with low recurrence which could be treated with lumpectomy without postoperative radiation treatment. While a number of attempts have been made to classify early stage breast cancer, no consensus recommendation on postoperative radiation treatment has been obtained from these studies. Page *et al.*, Cancer 75:1219 (1995); Fisher *et al.*, Cancer 75:1223 (1995); Silverstein *et al.*, Cancer 77:2267 (1996).

As discussed above, each of the methods for diagnosing and staging ovarian, pancreatic or breast cancer is limited by the technology employed. Accordingly, there is need for sensitive molecular and cellular markers for the detection of ovarian, pancreatic or breast cancer. There is a need for molecular markers for the accurate staging, including clinical and pathological staging, of ovarian, pancreatic or breast cancers to optimize treatment methods. Finally, there is a need for sensitive molecular and cellular markers to monitor the progress of cancer treatments, including markers that can detect recurrence of ovarian, pancreatic or breast cancers following remission.

The present invention provides alternative methods of treating ovarian, pancreatic or breast cancer that overcome the limitations of conventional therapeutic methods as well as offer additional advantages that will be apparent from the detailed description below.

SUMMARY OF THE INVENTION

The present invention solves many needs in the art by providing nucleic acid molecules, polypeptides and antibodies thereto, variants and derivatives of the nucleic acids and polypeptides, agonists and antagonists that may be used to identify, diagnose, monitor, stage, image and treat ovarian cancer and non-cancerous disease states in ovarian; identify and monitor ovarian tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered ovarian tissue for treatment and research.

One aspect of the present invention relates to nucleic acid molecules that are specific to ovarian cells, ovarian tissue and/or the ovarian organ. These ovarian specific nucleic acids (OSNAs) may be a naturally occurring cDNA, genomic DNA, RNA, or a

fragment of one of these nucleic acids, or may be a non-naturally occurring nucleic acid molecule. If the OSNA is genomic DNA, then the OSNA is a ovarian specific gene (OSG). If the OSNA is RNA, then it is a ovarian specific transcript encoded by a OSG. Due to alternative splicing and transcriptional modification one OSG may encode for

5 multiple ovarian specific RNAs. In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to ovarian. More preferred is a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 9-14. In another preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1-8. For the OSNA sequences listed herein,

10 DEX0439_001.nt.1 corresponds to SEQ ID NO: 1. For sequences with multiple splice variants, the parent sequence DEX0439_001.nt.1, will be followed by DEX0439_001.nt.2, etc. for each splice variant. The sequences off the corresponding peptides are listed as DEX0439_001.aa.1, etc. For the mapping of all of the nucleotides and peptides, see the table in the Example 1 section below.

15 This aspect of the present invention also relates to nucleic acid molecules that selectively hybridize or exhibit substantial sequence similarity to nucleic acid molecules encoding a ovarian Specific Protein (OSP), or that selectively hybridize or exhibit substantial sequence similarity to a OSNA. In one embodiment of the present invention the nucleic acid molecule comprises an allelic variant of a nucleic acid molecule encoding

20 a OSP, or an allelic variant of a OSNA. In another embodiment, the nucleic acid molecule comprises a part of a nucleic acid sequence that encodes a OSP or a part of a nucleic acid sequence of a OSNA.

In addition, this aspect of the present invention relates to a nucleic acid molecule further comprising one or more expression control sequences controlling the transcription

25 and/or translation of all or a part of a OSNA or the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of a OSP.

Another aspect of the present invention relates to vectors and/or host cells comprising a nucleic acid molecule of this invention. In a preferred embodiment, the nucleic acid molecule of the vector and/or host cell encodes all or a fragment of a OSP. In

30 another preferred embodiment, the nucleic acid molecule of the vector and/or host cell comprises all or a part of a OSNA. Vectors and host cells of the present invention are useful in the recombinant production of polypeptides, particularly OSPs of the present invention.

Another aspect of the present invention relates to polypeptides encoded by a nucleic acid molecule of this invention. The polypeptide may comprise either a fragment or a full-length protein. In a preferred embodiment, the polypeptide is a OSP. However, this aspect of the present invention also relates to mutant proteins (muteins) of OSPs, fusion proteins of which a portion is a OSP, and proteins and polypeptides encoded by allelic variants of a OSNA as provided herein.

A further aspect of the present invention is a novel splice variant which encodes an amino acid sequence that provides a novel region to be targeted for the generation of reagents that can be used in the detection and/or treatment of cancer. The novel amino acid sequence may lead to a unique protein structure, protein subcellular localization, biochemical processing or function. This information can be used to directly or indirectly facilitate the generation of additional or novel therapeutics or diagnostics. The nucleotide sequence in this novel splice variant can be used as a nucleic acid probe for the diagnosis and/or treatment of cancer.

Another aspect of the present invention relates to antibodies and other binders that specifically bind to a polypeptide of the instant invention. Accordingly antibodies or binders of the present invention specifically bind to OSPs, muteins, fusion proteins, and/or homologous proteins or polypeptides encoded by allelic variants of an OSNA as provided herein.

Another aspect of the present invention relates to agonists and antagonists of the nucleic acid molecules and polypeptides of this invention. The agonists and antagonists of the instant invention may be used to treat ovarian cancer and non-cancerous disease states in ovarian and to produce engineered ovarian tissue.

Another aspect of the present invention relates to methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. Such methods are useful in identifying, diagnosing, monitoring, staging, imaging and treating ovarian cancer and non-cancerous disease states in ovarian. Such methods are also useful in identifying and/or monitoring ovarian tissue. In addition, measurement of levels of one or more of the nucleic acid molecules of this invention may be useful for diagnostics as part of panel in combination with known other markers, particularly those described in the ovarian cancer background section above.

Another aspect of the present invention relates to use of the nucleic acid molecules of this invention in gene therapy, for producing transgenic animals and cells, and for producing engineered ovarian tissue for treatment and research.

Another aspect of the present invention relates to methods for detecting
5 polypeptides this invention, preferably using antibodies thereto. Such methods are useful to identify, diagnose, monitor, stage, image and treat ovarian cancer and non-cancerous disease states in ovarian. In addition, measurement of levels of one or more of the polypeptides of this invention may be useful to identify, diagnose, monitor, stage, image ovarian cancer in combination with known other markers, particularly those described in
10 the ovarian cancer background section above. The polypeptides of the present invention can also be used to identify and/or monitor ovarian tissue, and to produce engineered ovarian tissue.

Yet another aspect of the present invention relates to a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the
15 computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences. In addition, the computer records regarding the nucleic acid and/or amino acid sequences and/or measurements of their levels may be used alone or in combination with other markers to diagnose ovarian related diseases.

20 BREIF DESCRIPTION OF THE FIGURES

FIGURE 1 is a nucleotide sequence alignment which shows regions of similarity and difference between DEX0439_001.nt.1 (Ovr107) and DEX0439_001.nt.2;

FIGURE 2 is an amino acid sequence alignment which shows regions of similarity and difference between DEX0439_001.aa.1 (Ovr107) and DEX0439_001.aa.2;

25 FIGURE 3 is an amino acid sequence alignment which shows regions of similarity and difference between DEX0439_001.aa.1 (Ovr107) and DEX0439_001.aa.3;

FIGURE 4 is a nucleotide sequence alignment which shows regions of similarity and difference between DEX0439_001.nt.1 (Ovr107) and DEX0439_00X.nt.4 (Ovr107v4);

30 FIGURE 5 is a nucleotide sequence alignment which shows regions of similarity and difference between DEX0439_001.nt.1 (Ovr107) and DEX0439_001.nt.6;

FIGURE 6 is a nucleotide sequence alignment which shows regions of similarity and difference between DEX0439_001.nt.1 (Ovr107) and DEX0439_001.nt.8;

FIGURE 7 is a nucleotide sequence alignment which shows regions of similarity and difference between DEX0439_001.nt.1 (Ovr107) and DEX0439_001.nt.10;

5 FIGURE 8 is a nucleotide sequence alignment which shows regions of similarity and difference between DEX0439_002.nt.1 (Ovr110) and DEX0439_002.nt.2 (Ovr110v1);

FIGURE 9 is an amino acid sequence alignment which shows regions of similarity and difference between DEX0439_002.aa.1 (Ovr110) and DEX0439_002.aa.2;

10 FIGURE 10 is an amino acid sequence alignment which shows regions of similarity and difference between DEX0439_002.aa.1 (Ovr110) and DEX0439_002.aa.3.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection
15 with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid
20 chemistry and hybridization described herein are those well known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g., Sambrook et al., Molecular Cloning: A*
25 *Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 3d ed., Cold Spring Harbor Press (2001); Ausubel et al., *Current Protocols in Molecular Biology*, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel et al., *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4th Ed.*, Wiley &
30 Sons (1999); Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, *Using Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press (1999).

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

- 10 A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and
- 15 "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single and double stranded forms of DNA. In addition, a polynucleotide may include either or both naturally occurring and modified nucleotides linked together by naturally occurring and/or non-naturally occurring nucleotide linkages.
- 20 Nucleotides are represented by single letter symbols in nucleic acid molecule sequences. The following table lists symbols identifying nucleotides or groups of nucleotides which may occupy the symbol position on a nucleic acid molecule. See Nomenclature Committee of the International Union of Biochemistry (NC-IUB), Nomenclature for incompletely specified bases in nucleic acid sequences,
- 25 Recommendations 1984., *Eur J Biochem.* 150(1):1-5 (1985).

Symbol	Meaning	Group/Origin of Designation	Complementary Symbol
a	a	Adenine	t/u
g	g	Guanine	c
c	c	Cytosine	g
t	t	Thymine	a
u	u	Uracil	a
r	g or a	puRine	y
y	t/u or c	pYrimidine	r
m	a or c	aMino	k
k	g or t/u	Keto	m
s	g or c	Strong interactions 3H-bonds	w
w	a or t/u	Weak interactions 2H-bonds	s
b	g or c or t/u	not a	v

d	a or g or t/u	not c	h
h	a or c or t/u	not g	d
v	a or g or c	not t, not u	b
n	a or g or c or t/u, unknown, or other	any	n

The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

5 An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its
10 naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in
15 reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as
20 integrated into a host cell chromosome.

A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to
25 occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or
30 synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and U.S. Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at

least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single-or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including in vitro recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

The term "naturally occurring nucleotide" referred to herein includes naturally occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides"

referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phosphoraniladate, phosphoroamidate, and the like. *See e.g.*,
5 LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), and U.S. Patent No. 5,151,510, the disclosure of which is hereby incorporated by
10 reference in its entirety.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred
15 to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

20 The term "allelic variant" refers to one of two or more alternative naturally occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence. In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences
25 refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number
30 of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2

and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998)). Unless otherwise

5 specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1.

A reference to a nucleic acid sequence encompasses its complement unless
10 otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, double stranded RNA (dsRNA) inhibition (RNAi), combination of triplex and antisense, hybridization probes and PCR primers.

15 In the molecular biology art, researchers use the terms “percent sequence identity”, “percent sequence similarity” and “percent sequence homology” interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term “substantial similarity” or “substantial sequence similarity,” when
20 referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of
25 the nucleotide bases, as measured by any well known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists between a first and second nucleic acid sequence when the first nucleic acid sequence or fragment thereof hybridizes to an antisense strand of the second nucleic acid, under selective hybridization conditions.
30 Typically, selective hybridization will occur between the first nucleic acid sequence and an antisense strand of the second nucleic acid sequence when there is at least about 55% sequence identity between the first and second nucleic acid sequences— preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% —

over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. See Sambrook (1989), *supra*, p. 9.51.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$$T_m = 81.5^\circ\text{C} + 16.6 (\log_{10}[\text{Na}^+]) + 0.41 (\text{fraction G} + \text{C}) - 0.63 (\% \text{ formamide}) - (600/l) \text{ where } l \text{ is the length of the hybrid in base pairs.}$$

The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.35 (\% \text{ formamide}) - (820/l).$$

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.50 (\% \text{ formamide}) - (820/l).$$

In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also

hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well known in the art.

An example of stringent hybridization conditions for hybridization of
5 complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours and preferably overnight. An example of moderate stringency
10 hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or northern blot or for screening a library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid
15 sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (e.g. 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These
20 agents are well known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook
25 (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in
30 the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acids that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid is

created synthetically or recombinantly using a high codon degeneracy as permitted by the redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (*e.g.*, for oligonucleotide probes) may be calculated by the formula:

5 $T_m = 81.5^{\circ}\text{C} + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\text{fraction G+C}) - (600/\text{N})$, wherein N is
change length and the $[\text{Na}^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For
hybridization of probes shorter than 100 nucleotides, hybridization is usually performed
under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0
pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched
10 probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and
methods for empirically determining hybridization conditions are well known in the art.
See, e.g., Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the
DNA with a restriction enzyme that acts only at certain sequences in the DNA. The
15 various restriction enzymes referred to herein are commercially available and their
reaction conditions, cofactors and other requirements for use are known and routine to the
skilled artisan. For analytical purposes, typically, 1 µg of plasmid or DNA fragment is
digested with about 2 units of enzyme in about 20 µl of reaction buffer. For the purpose of
isolating DNA fragments for plasmid construction, typically 5 to 50 µg of DNA are
20 digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate
buffers and substrate amounts for particular restriction enzymes are described in standard
laboratory manuals, such as those referenced below, and are specified by commercial
suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions
may vary in accordance with standard procedures, the supplier's instructions and the
25 particulars of the reaction. After digestion, reactions may be analyzed, and fragments may
be purified by electrophoresis through an agarose or polyacrylamide gel, using well
known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between
two or more polynucleotides, which most often are double-stranded DNAs. Techniques
30 for ligation are well known to the art and protocols for ligation are described in standard
laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.

Genome-derived "single exon probes," are probes that comprise at least part of an
exon ("reference exon") and can hybridize detectably under high stringency conditions to

transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies. In another aspect, the invention is directed to single exon probes based on the OSNAs disclosed herein.

In one embodiment, the term "microarray" refers to a "nucleic acid microarray" having a substrate-bound plurality of nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). Additionally, these nucleic acid microarrays include substrate-bound plurality of nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000). Examples of nucleic acid microarrays may be found in U.S. Patent Nos. 6,391,623, 6,383,754, 6,383,749, 6,380,377, 6,379,897, 6,376,191, 6,372,431, 6,351,712, 6,344,316, 6,316,193, 6,312,906, 6,309,828, 6,309,824, 6,306,643, 6,300,063, 6,287,850, 6,284,497, 6,284,465, 6,280,954, 6,262,216, 6,251,601, 6,245,518, 6,263,287, 6,251,601, 6,238,866, 6,228,575, 6,214,587, 6,203,989, 6,171,797, 6,103,474, 6,083,726, 6,054,274, 6,040,138, 6,083,726, 6,004,755, 6,001,309, 5,958,342, 5,952,180, 5,936,731, 5,843,655, 5,814,454, 5,837,196, 5,436,327, 5,412,087, 5,405,783, the disclosures of which are incorporated herein by reference in their entireties.

In an alternative embodiment, a "microarray" may also refer to a "peptide microarray" or "protein microarray" having a substrate-bound collection of plurality of polypeptides, the binding to each of the plurality of bound polypeptides being separately detectable. Alternatively, the peptide microarray may have a plurality of binders, including but not limited to monoclonal antibodies, polyclonal antibodies, phage display binders, yeast 2 hybrid binders, aptamers, which can specifically detect the binding of the polypeptides of this invention. The array may be based on autoantibody detection to the polypeptides of this invention, see Robinson *et al.*, *Nature Medicine* 8(3):295-301 (2002). Examples of peptide arrays may be found in WO 02/31463, WO 02/25288, WO 01/94946, WO 01/88162, WO 01/68671, WO 01/57259, WO 00/61806, WO 00/54046, WO 00/47774, WO 99/40434, WO 99/39210, WO 97/42507 and U.S. Patent Nos. 6,268,210, 5,766,960, 5,143,854, the disclosures of which are incorporated herein by reference in their entireties.

In addition, determination of the levels of the OSNA or OSP may be made in a multiplex manner using techniques described in WO 02/29109, WO 02/24959, WO 01/83502, WO01/73113, WO 01/59432, WO 01/57269, WO 99/67641, the disclosures of which are incorporated herein by reference in their entireties.

The term "mutant", "mutated", or "mutation" when applied to nucleic acid sequences means that nucleotides in a nucleic acid sequence may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment of the present invention, the nucleic acid sequence is the wild type nucleic acid sequence encoding a OSP or is a OSNA. The nucleic acid sequence may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. See, e.g., Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g., Reidhaar-Olson et al., Science* 241: 53-57 (1988).

5 The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

10 The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g., Stemmer, Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

15 The term "*in vivo* mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

20 The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

25 The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. *See, e.g., Arkin et al., Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

30 The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. *See, e.g., Delegrave et al., Biotechnology*

Research 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993).

“Operatively linked” expression control sequences refers to a linkage in which the expression control sequence is either contiguous with the gene of interest to control the gene of interest, or acts in *trans* or at a distance to control the gene of interest.

The term “expression control sequence” as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences.

Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence. The term “control sequences” is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

The term “vector,” as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a “plasmid”, which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as “recombinant expression vectors” (or simply, “expression vectors”).

In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that serve equivalent
5 functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which a recombinant expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in
10 succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refers to that portion of a transcript-derived nucleic acid that can be translated in its
15 entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

20 As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence is meant to be inclusive of all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally occurring and non-naturally
25 occurring proteins and polypeptides, as well as polypeptide fragments and polypeptide mutants, derivatives and analogs thereof. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises a OSP encoded by a nucleic acid molecule of the
30 instant invention, or a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins

from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered
5 substantially free of naturally associated components by isolation, using protein purification techniques well known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A
10 substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be determined by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel with a stain well
15 known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well known in the art for purification.

The term "fragment" when used herein with respect to polypeptides of the present invention refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length OSP. In a preferred embodiment, the fragment is a
20 contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally occurring polypeptide. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even
25 more preferably at least 70 amino acids long.

A "derivative" when used herein with respect to polypeptides of the present invention refers to a polypeptide which is substantially similar in primary structural sequence to a OSP but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the OSP. Such modifications include, for example,
30 acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation,

formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , ^{14}C and ^3H , ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well known in the art. *See Ausubel (1992), supra;* Ausubel (1999), *supra*.

The term "fusion protein" refers to polypeptides of the present invention coupled to a heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence that encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally occurring sequence.

Analogues typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: $-\text{CH}_2\text{NH}-$, $-\text{CH}_2\text{S}-$, $-\text{CH}_2\text{-CH}_2-$, $-\text{CH}=\text{CH}-$ (*cis* and *trans*), $-\text{COCH}_2-$, $-\text{CH}(\text{OH})\text{CH}_2-$, and $-\text{CH}_2\text{SO}-$, by methods well known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al.*, *Ann. Rev. Biochem.* 61:387-418 (1992)). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

The term "mutant" or "mutein" when referring to a polypeptide of the present invention relates to an amino acid sequence containing substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a OSP. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to a OSP. In an even more preferred embodiment, a mutein exhibits

95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as GAP or BESTFIT or other variation Smith-Waterman alignment. See, T. F. Smith and M. S. Waterman, J. Mol. Biol. 147:195-197 (1981) and
5 W.R. Pearson, Genomics 11:635-650 (1991).

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example,
10 single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions
15 are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (e.g., a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are
20 described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991).

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. See Golub *et al.* (eds.), Immunology - A Synthesis 2nd Ed., Sinauer
25 Associates (1991). Stereoisomers (e.g., D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α -, α -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include:
4-hydroxyproline, γ -carboxyglutamate, ϵ -N,N,N-trimethyllysine, ϵ -N-acetyllysine,
30 O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (e.g., 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the

amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

By "homology" or "homologous" when referring to a polypeptide of the present invention it is meant polypeptides from different organisms with a similar sequence to the encoded amino acid sequence of a OSP and a similar biological activity or function. Although two polypeptides are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the polypeptides. Instead, the term "homologous" is defined to mean that the two polypeptides have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous polypeptide is one that exhibits 50% sequence similarity to OSP, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are homologous polypeptides that exhibit 80%, 85% or 90% sequence similarity to a OSP. In a yet more preferred embodiment, a homologous polypeptide exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to polypeptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. See, e.g., Pearson, *Methods Mol. Biol.* 24: 307-31 (1994).

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Serine (S), Threonine (T);
- 2) Aspartic Acid (D), Glutamic Acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);

- 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992). A "moderately conservative" replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutin thereof. *See, e.g.*, GCG Version 6.1. Other programs include FASTA, discussed *supra*.

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. *See, e.g.*, Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997). Preferred parameters for blastp are:

Expectation value:	10 (default)
Filter:	seg (default)
Cost to open a gap:	11 (default)
Cost to extend a gap:	1 (default)
Max. alignments:	100 (default)
Word size:	11 (default)
No. of descriptions:	100 (default)
Penalty Matrix:	BLOSUM62

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

Algorithms other than blastp for database searching using amino acid sequences are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*). For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1.

An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', F(ab')₂, Fv, dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. A Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; a F(ab')₂ fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; a Fd fragment consists of the VH and CH1 domains; a Fv fragment consists of the VL and VH domains of a single arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g.*, Ward *et al.*, *Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" as used herein it is meant the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which VL and VH regions are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g.*, Bird *et al.*, *Science* 242: 423-426 (1988); Huston *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain

and creating two antigen binding sites. See e.g., Holliger *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al.*, *Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

10 An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

15 An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (e.g., BSA) or a chemical such as polyethylene glycol (PEG).

20 A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

25 The term "epitope" includes any protein determinant capable of specific binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than 1 μ M, preferably less than 100 nM and most preferably less than 10 nM.

30 The term "patient" includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term "ovarian specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the ovarian as compared to other tissues in the body. In a preferred embodiment, a "ovarian specific" nucleic acid molecule or polypeptide is detected at a level that is 1.5-fold higher than any other tissue in the body. In a more preferred embodiment, the "ovarian specific" nucleic acid molecule or polypeptide is detected at a level that is 2-fold higher than any other tissue in the body, more preferably 5-fold higher, still more preferably at least 10-fold, 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels, such as Western blot analysis.

15 Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

One aspect of the invention provides isolated nucleic acid molecules that are specific to the ovarian or to ovarian cells or tissue or that are derived from such nucleic acid molecules. These isolated ovarian specific nucleic acids (OSNAs) may comprise cDNA genomic DNA, RNA, or a combination thereof, a fragment of one of these nucleic acids, or may be a non-naturally occurring nucleic acid molecule. A OSNA may be derived from an animal. In a preferred embodiment, the OSNA is derived from a human or other mammal. In a more preferred embodiment, the OSNA is derived from a human or other primate. In an even more preferred embodiment, the OSNA is derived from a human.

In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to ovarian, a ovarian-specific polypeptide (OSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 9-14. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1-8. Nucleotide sequences of the instantly-described nucleic acid molecules were determined by

assembling several DNA molecules from either public or proprietary databases. Some of the underlying DNA sequences are the result, directly or indirectly, of at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Amersham Biosciences, Sunnyvale, CA, USA).

Nucleic acid molecules of the present invention may also comprise sequences that selectively hybridizes to a nucleic acid molecule encoding a OSNA or a complement or antisense thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may or may not encode a OSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes a OSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 9-14. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1-8 or the antisense sequence thereof. Preferably, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a OSP under low stringency conditions. More preferably, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a OSP under moderate stringency conditions. Most preferably, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a OSP under high stringency conditions. In a preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 9-14. In a more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1-8.

Nucleic acid molecules of the present invention may also comprise nucleic acid sequences that exhibit substantial sequence similarity to a nucleic acid encoding a OSP or a complement of the encoding nucleic acid molecule. In this embodiment, it is preferred that the nucleic acid molecule exhibit substantial sequence similarity to a nucleic acid

molecule encoding human OSP. More preferred is a nucleic acid molecule exhibiting substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 9-14. By substantial sequence similarity it is meant a nucleic acid molecule having at least 60% sequence identity with a nucleic acid molecule encoding a OSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 9-14, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding a OSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. Most preferred in this embodiment is a nucleic acid molecule that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding a OSP.

The nucleic acid molecules of the present invention are also inclusive of those exhibiting substantial sequence similarity to a OSNA or its complement. In this embodiment, it is preferred that the nucleic acid molecule exhibit substantial sequence similarity to a nucleic acid molecule having a nucleic acid sequence of SEQ ID NO: 1-8. By substantial sequence similarity it is meant a nucleic acid molecule that has at least 60% sequence identity with a OSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1-8, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. More preferred is a nucleic acid molecule that has at least 90% sequence identity with a OSNA, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. Most preferred is a nucleic acid molecule that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a OSNA.

Nucleic acid molecules that exhibit substantial sequence similarity are inclusive of sequences that exhibit sequence identity over their entire length to a OSNA or to a nucleic acid molecule encoding a OSP, as well as sequences that are similar over only a part of its length. In this case, the part is at least 50 nucleotides of the OSNA or the nucleic acid molecule encoding a OSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally occurring one that is derived from another species, especially one derived from another primate, wherein

the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 9-14 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO: 1-8. The similar nucleic acid molecule may also be a naturally occurring nucleic acid molecule from a human, when the OSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed mutation of a OSNA. In a preferred embodiment, the substantially similar nucleic acid molecule is an OSNA.

The nucleic acid molecules of the present invention are also inclusive of allelic variants of a OSNA or a nucleic acid encoding a OSP. For example, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes and the sequence determined from one individual of a species may differ from other allelic forms present within the population. More than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001) – Variants with small deletions and insertions of more than a single nucleotide are also found in the general population, and often do not alter the function of the protein. In addition, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes a OSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes a OSP comprising an amino acid sequence of SEQ ID NO: 9-14. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is a OSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the

nucleic acid sequence of SEQ ID NO: 1-8. Also preferred is that the allelic variant is a naturally occurring allelic variant in the species of interest, particularly human.

Nucleic acid molecules of the present invention are also inclusive of nucleic acid sequences comprising a part of a nucleic acid sequence of the instant invention. The part
5 may or may not encode a polypeptide, and may or may not encode a polypeptide that is a OSP. In a preferred embodiment, the part encodes a OSP. In one embodiment, the nucleic acid molecule comprises a part of a OSNA. In another embodiment, the nucleic acid molecule comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to a OSNA. In another embodiment, the nucleic acid
10 molecule comprises a part of a nucleic acid molecule that is an allelic variant of a OSNA. In yet another embodiment, the nucleic acid molecule comprises a part of a nucleic acid molecule that encodes a OSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than
15 the sequence of the nucleic acid molecule encoding the full-length protein.

Nucleic acid molecules of the present invention are also inclusive of nucleic acid sequences that encode fusion proteins, homologous proteins, polypeptide fragments, muteins and polypeptide analogs, as described *infra*.

Nucleic acid molecules of the present invention are also inclusive of nucleic acid
20 sequences containing modifications of the native nucleic acid molecule. Examples of such modifications include, but are not limited to, nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that may be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is
25 used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent,
30 the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

Accordingly, in one embodiment, a nucleic acid molecule may include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or

fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. The labeled nucleic acid molecules are particularly useful as hybridization probes.

Common radiolabeled analogues include those labeled with ^{33}P , ^{32}P , and ^{35}S , such as α - ^{32}P -dATP, α - ^{32}P -dCTP, α - ^{32}P -dGTP, α - ^{32}P -dTTP, α - ^{32}P -3'dATP, α - ^{32}P -ATP, α - ^{32}P -CTP, α - ^{32}P -GTP, α - ^{32}P -UTP, α - ^{35}S -dATP, γ - ^{35}S -GTP, γ - ^{33}P -dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Biosciences, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. See Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000).

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules of the present invention can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase

chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and Peptide Nucleic Acids (PNA) to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); *see Alers et al., Genes, Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994); Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994). Alternatively, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. *See, e.g.*, Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids Symp. Ser.* (37): 255-6 (1997); and U.S. Patent Nos. 5,846,726, 5,925,517, 5,925,517, 5,723,591 and 5,538,848, the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the present invention may also be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. *See Hartmann et al. (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science*, Kluwer Law International (1999); Stein *et al. (eds.), Applied Antisense Oligonucleotide Technology*, Wiley-Liss (1998); Chadwick *et al. (eds.), Oligonucleotides as Therapeutic Agents – Symposium No. 209*, John Wiley & Son

Ltd (1997). Such altered internucleoside bonds are often desired for techniques or for targeted gene correction, Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000). For double stranded RNA inhibition which may utilize either natural ds RNA or ds RNA modified in its, sugar, phosphate or base, see Hannon, *Nature* 418(11): 244-251 (2002);
5 Fire *et al.* in WO 99/32619; Tuschl *et al.* in US2002/0086356; Kruetzer *et al.* in WO 00/44895, the disclosures of which are incorporated herein by reference in their entirety;. For circular antisense, see Kool in U.S. Patent No. 5,426,180, the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation,
10 phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having
15 normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative U.S. Patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U.S. Patent Nos. 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131;
20 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but
25 have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and
30 thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts. Representative U.S. patents that teach the

preparation of the above backbones include, but are not limited to, U.S. Patent Nos. 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred nucleic acid molecules, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Patent Nos. 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference in its entirety. Automated PNA synthesis is readily achievable on commercial synthesizers (*see, e.g.*, "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA). PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The T_m of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the T_m of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the T_m by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the T_m by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyamide backbone with

nucleobase sidechains. *See, e.g., Ray et al., FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al., Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al., Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999).

5 Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in, Misra *et al., Biochem.* 37: 1917-1925
10 (1998); and Finn *et al., Nucl. Acids Res.* 24: 3357-3363 (1996), and U.S. Patent Nos. 5,760,012 and 5,731,181, the disclosures of which are incorporated herein by reference in their entireties.

 Unless otherwise specified, nucleic acid molecules of the present invention can include any topological conformation appropriate to the desired use; the term thus
15 explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al., Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al., Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); and Nilsson
20 *et al., Science* 265(5181): 2085-8 (1994). Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al., Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al., Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al., J. Mol. Med.* 75(4): 267-82 (1997); Rowley *et al., Mol Med* 5(10): 693-700 (1999); Kool, *Annu Rev Biophys Biomol Struct.* 25: 1-28 (1996).

25

Methods for Using Nucleic Acid Molecules as Probes and Primers

 The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid
30 samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect and characterize gross alterations in the gene of a OSNA, such as deletions, insertions, translocations, and duplications of the OSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g.,*

5 Andreeff *et al.* (eds.), Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999). The isolated nucleic acid molecules of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.,* Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic

10 clones that include a nucleic acid molecule of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level. Alternatively, detection techniques such as molecular beacons may be used, *see* Kostrikis *et al. Science* 279:1228-1229 (1998).

15 The isolated nucleic acid molecules of the present invention can be also be used as probes to detect, characterize, and quantify OSNA in, and isolate OSNA from, transcript-derived nucleic acid samples. In one embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A⁺- selected RNA samples. In

20 another embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g.,* Schwarchzacher *et al., In Situ Hybridization*, Springer-Verlag New York (2000). In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization

25 probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to OSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic

30 acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel

(1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000).

In another embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify and/or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In this embodiment, it is preferred that the probe or primer be derived from a nucleic acid molecule encoding a OSP. More preferably, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 9-14. Also preferred are probes or primers derived from a OSNA. More preferred are probes or primers derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-8.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well known in the art. See, e.g., Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); and McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995). Methods for performing RT-PCR are collected, e.g., in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR,

Eaton Publishing Company/Bio Techniques Books Division, 1998; and Siebert (ed.), PCR Technique:RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995).

5 PCR and hybridization methods may be used to identify and/or isolate nucleic acid molecules of the present invention including allelic variants, homologous nucleic acid molecules and fragments. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules of the present invention that encode homologous proteins, analogs, fusion protein or muteins of the invention. Nucleic acid primers as described herein can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

10 These nucleic acid primers can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Pat. No. 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); international patent publications WO 97/19193 and WO 00/15779, and U.S. Patent
15 Nos. 5,854,033 and 5,714,320, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

20 Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

25 In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or positively charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another
30 embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene,

polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

5 The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization
10 to each of the plurality of bound nucleic acids being separately detectable. At low density, e.g. on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect
15 of the invention to provide microarrays that comprise one or more of the nucleic acid molecules of the present invention.

In yet another embodiment, the invention is directed to single exon probes based on the OSNAs disclosed herein.

20 *Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides*

Another aspect of the present invention provides vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

25 The vectors can be used, *inter alia*, for propagating the nucleic acid molecules of the present invention in host cells (cloning vectors), for shuttling the nucleic acid molecules of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acid molecules of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of
30 the nucleic acid molecules of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acid molecules of the present invention, alone or as fusion proteins with heterologous polypeptides (expression vectors). Vectors are by now well known in the art, and are described, *inter alia*, in Jones *et al.* (eds.),

Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*. Furthermore, a variety of vectors are available commercially. Use of existing vectors and modifications thereof are well within the skill in the art. Thus, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences that control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989, λ GT10 and λ GT11, and other phages, *e.g.*, M13 and filamentous single stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline,

chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEplac series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2 μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as *ura3-52*, *his3-D1*, *leu2-D1*, *trp1-D1* and *lys2-201*.

Insect cells may be chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda*, *e.g.*, Sf9 and Sf21 cell lines, and expresSFTM cells (Protein Sciences Corp., Meriden, CT; USA), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

The host cells may also be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral

origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and
5 thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include, include but are not
10 limited to, resistance to neomycin (G418), blasticidin, hygromycin and zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and
15 retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

20 It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention
25 may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the nucleic acid molecules of this invention. Such useful expression control sequences include the expression control sequences associated with structural
30 genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the

transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that
5 modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the *trc* promoter, a hybrid derived from the *trp* and *lac* promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC
10 system, the major operator and promoter regions of phage lambda, the control regions of *fd* coat protein, and the *araBAD* operon. Prokaryotic expression vectors may further include transcription terminators, such as the *aspA* terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

15 Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the *CYC1* promoter, the *GAL1* promoter, the *GAL10* promoter, *ADH1* promoter, the promoters of the yeast α -mating system, or the *GPD* promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the *CYC1* or *ADH1* gene.

20 Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include, but are not limited to, those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-
25 promoter from SV40 and the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the OSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the
30 polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β -globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows a high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the *trc* promoter, which is regulated by the *lac* operon, and the *pL* promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid *Plac/ara-1* promoter and the *PLtetO-1* promoter. The *PLtetO-1* promoter takes advantage of the high expression levels from the *PL* promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the *Tn10* tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one embodiment of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Such tags include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc.,

Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverly, MA, USA).

- 5 Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the polypeptides of the present
- 10 invention can be expressed as a fusion to glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a
- 15 myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope, detectable by anti-HA antibody.

- For secretion of expressed polypeptides, vectors can include appropriate sequences
- 20 that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

- Expression vectors can also be designed to fuse proteins encoded by the
- 25 heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful protein fusions include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusions for use in two hybrid systems.

- 30 Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. See Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins:

A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the α -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See Li et al., J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well known in the art. *See Heim et al., Curr. Biol.* 6: 178-182 (1996) and Palm *et al., Methods Enzymol.* 302: 378-394 (1999). A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g. Cormack et al., Gene* 173: 33-38 (1996); U.S. Patent Nos. 6,090,919 and 5,804,387, the disclosures of which are incorporated herein by reference in their entireties) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g. Heim et al., Curr. Biol.* 6: 178-182 (1996) and Cormack *et al., Gene*

173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g., Heim et al., Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al., Nature* 388: 882-887 (1997)) and Citrine (*see, e.g., Heikal et al., Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patent Nos. 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. *See also Conn (ed.), Green Fluorescent Protein (Methods in Enzymology,* 10 *Vol. 302), Academic Press, Inc. (1999); Yang, et al., J Biol Chem,* 273: 8212-6 (1998); Bevis *et al., Nature Biotechnology,* 20:83-7 (2002). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half-life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412, the disclosures of which are incorporated herein by reference in their entireties.

For long-term, high-yield recombinant production of the polypeptides of the present invention, stable expression is preferred. Stable expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo

Alto, CA, USA) allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid molecules of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed polypeptide in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation, and it is an aspect of the present invention to provide OSPs with such post-translational modifications.

In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid molecules of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid molecules of this invention.

The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed

from the nucleic acid molecules according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell (e.g., conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well known in the art (See, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda* — e.g., Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA) — *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2

cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (e.g., from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well known and readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National
5 Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from ovarian are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human ovarian cells.

Particular details of the transfection, expression and purification of recombinant
10 proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. See, e.g., Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*.

15 Methods for introducing the vectors and nucleic acid molecules of the present invention into the host cells are well known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be packaged
20 using a packaging extract (e.g., Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, e.g., with CaCl_2 , or a solution of Mg^{2+} , Mn^{2+} , Ca^{2+} , Rb^+ or K^+ , dimethyl
25 sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (e.g., Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5α competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent *E. coli* Kit
30 (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent to take up exogenous DNA by electroporation by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided by BioRad (Richmond, CA, USA).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action of hydrolytic enzymes such as a snail-gut extract, usually denoted Glusulase or Zymolyase, or an enzyme from *Arthrobacter luteus* to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+} . Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate to permeabilize the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with CaPO_4 or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for CaPO_4 transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent, FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found in, for example, ; Norton *et*

al. (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000). Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.*, Cheng *et al.*, *Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA* 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

Purification of recombinantly expressed proteins is now well within the skill in the art and thus need not be detailed here. *See, e.g.*, Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001).

Briefly, however, if purification tags have been fused through use of an expression vector that appends such tag, purification can be effected, at least in part, by means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

Polypeptides, including Fragments Muteins, Homologous Proteins, Allelic Variants, Analogs and Derivatives

Another aspect of the invention relates to polypeptides encoded by the nucleic acid molecules described herein. In a preferred embodiment, the polypeptide is an ovarian specific polypeptide (OSP). In an even more preferred embodiment, the polypeptide comprises an amino acid sequence of SEQ ID NO:9-14 or is derived from a polypeptide having the amino acid sequence of SEQ ID NO: 9-14. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well known to those having ordinary skill in the art.

Polypeptides of the present invention may also comprise a part or fragment of a OSP. In a preferred embodiment, the fragment is derived from a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO: 9-14.

Polypeptides of the present invention comprising a part or fragment of an entire OSP may
5 or may not be OSPs. For example, a full-length polypeptide may be ovarian-specific, while a fragment thereof may be found in other tissues as well as in ovarian. A polypeptide that is not a OSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-OSP antibodies. In a preferred embodiment, the part or fragment is a OSP. Methods
10 of determining whether a polypeptide of the present invention is a OSP are described *infra*.

Polypeptides of the present invention comprising fragments of at least 6 contiguous amino acids are also useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g., Geysen et al., Proc. Natl. Acad. Sci. USA* 81: 3998-4002
15 (1984) and U.S. Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of a polypeptide of the present invention have utility in such a study.

20 Polypeptides of the present invention comprising fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize polypeptides of the present invention. *See, e.g., Lerner, Nature* 299: 592-596 (1982); Shinnick *et al., Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al., Science* 219: 660-6 (1983). As further described in the
25 above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic and are capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the polypeptides of the present invention have utility as immunogens.

Polypeptides comprising fragments of at least 8, 9, 10 or 12 contiguous amino
30 acids are also useful as competitive inhibitors of binding of the entire polypeptide, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind

specifically to the polypeptide of interest. See U.S. Patent Nos. 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

The polypeptide of the present invention thus preferably is at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the polypeptide of the present invention is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger polypeptides having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments by truncating the nucleic acid molecule, *e.g.*, a OSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally occurring polypeptide. Methods of producing polypeptide fragments are well known in the art. *See, e.g.*, Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a fragment, preferably a fragment of a OSP, may be produced by chemical or enzymatic cleavage of a OSP polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule of the present invention encoding a fragment, preferably of a OSP, in a host cell.

Polypeptides of the present invention are also inclusive of mutants, fusion proteins, homologous proteins and allelic variants.

A mutant protein, or mutein, may have the same or different properties compared to a naturally occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native polypeptide. Small deletions and insertions can often be found that do not alter the function of a protein. Muteins may or may not be ovarian-specific. Preferably, the mutein is ovarian-specific. More preferably the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 9-14. Accordingly, in a preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a OSP comprising an amino acid sequence of

SEQ ID NO: 9-14. In a yet more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to a OSP comprising an amino acid sequence of SEQ ID NO: 9-14.

5 A mutein may be produced by isolation from a naturally occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a
10 preferred embodiment, a mutein is produced from a host cell comprising a mutated nucleic acid molecule compared to the naturally occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid molecule of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino acids are altered, or may be
15 untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is ovarian-specific, as described below. Multiple random mutations can be introduced into the gene by methods well known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed
20 mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well known in the art. *See, e.g.*, Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), as well as U.S. Patent No.
25 5,223,408, which is herein incorporated by reference in its entirety.

 The invention also contemplates polypeptides that are homologous to a polypeptide of the invention. In a preferred embodiment, the polypeptide is homologous to a OSP. In an even more preferred embodiment, the polypeptide is homologous to a OSP selected from the group having an amino acid sequence of SEQ ID NO: 9-14. By
30 homologous polypeptide it is means one that exhibits significant sequence identity to a OSP, preferably a OSP having an amino acid sequence of SEQ ID NO: 9-14. By significant sequence identity it is meant that the homologous polypeptide exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more

- preferably at least 70%, yet more preferably at least 80% sequence identity to a OSP comprising an amino acid sequence of SEQ ID NO: 9-14. More preferred are homologous polypeptides exhibiting at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to a OSP comprising
- 5 an amino acid sequence of SEQ ID NO: 9-14. Most preferably, the homologous polypeptide exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to a OSP comprising an amino acid sequence of SEQ ID NO: 9-14. In a preferred embodiment, the amino acid substitutions of the homologous polypeptide are conservative amino acid substitutions as discussed above.
- 10 Homologous polypeptides of the present invention also comprise polypeptide encoded by a nucleic acid molecule that selectively hybridizes to a OSNA or an antisense sequence thereof. In this embodiment, it is preferred that the homologous polypeptide be encoded by a nucleic acid molecule that hybridizes to a OSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. More preferred is a
- 15 homologous polypeptide encoded by a nucleic acid sequence which hybridizes to a OSNA selected from the group consisting of SEQ ID NO: 1-8 or a homologous polypeptide encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes a OSP, preferably an OSP of SEQ ID NO:9-14 under low stringency, moderate stringency or high stringency conditions, as defined herein.
- 20 Homologous polypeptides of the present invention may be naturally occurring and derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, or baboon, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 9-14. The homologous polypeptide may also be a naturally occurring
- 25 polypeptide from a human, when the OSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally occurring polypeptide derived from a non-mammalian
- 30 species, such as birds or reptiles. The naturally occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. The homologous polypeptide may

also be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. Alternatively, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of a OSP. In a preferred
5 embodiment, the homologous polypeptide encodes a polypeptide that is a OSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated polypeptide not only identical in sequence to those described with particularity herein, but
10 also to provide isolated polypeptide ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in
15 eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, polypeptides of the present invention are also inclusive of those encoded by an allelic variant of a nucleic acid molecule encoding a OSP. In this embodiment, it is preferred that the polypeptide be encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid
20 sequence selected from the group consisting of SEQ ID NO: 9-14. More preferred is that the polypeptide be encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1-8.

Polypeptides of the present invention are also inclusive of derivative polypeptides encoded by a nucleic acid molecule according to the instant invention. In this
25 embodiment, it is preferred that the polypeptide be a OSP. Also preferred are derivative polypeptides having an amino acid sequence selected from the group consisting of SEQ ID NO: 9-14 and which has been acetylated, carboxylated, phosphorylated, glycosylated, ubiquitinated or other PTMs. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H . In another preferred
30 embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., www.expasy.org (accessed November 11, 2002), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications include, but are not limited to: (Z)-dehydrobutyrine; 1-chondroitin sulfate-L-aspartic acid ester; 1'-glycosyl-L-tryptophan; 1'-phospho-L-histidine; 1-thioglycine; 2'-(S-L-cysteinyl)-L-histidine; 2'-[3-carboxamido (trimethylammonio)propyl]-L-histidine; 2'-alpha-mannosyl-L-tryptophan; 2-methyl-L-glutamine; 2-oxobutanoic acid; 2-pyrrolidone carboxylic acid; 3'-(1'-L-histidyl)-L-tyrosine; 3'-(8alpha-FAD)-L-histidine; 3'-(S-L-cysteinyl)-L-tyrosine; 3', 3'', 5'-triiodo-L-thyronine; 3'-4'-phospho-L-tyrosine; 3-hydroxy-L-proline; 3'-methyl-L-histidine; 3-methyl-L-lanthionine; 3'-phospho-L-histidine; 4'-(L-tryptophan)-L-tryptophyl quinone; 42 N-cysteinyl-glycosylphosphatidylinositoethanolamine; 43 -(T-L-histidyl)-L-tyrosine; 4-hydroxy-L-arginine; 4-hydroxy-L-lysine; 4-hydroxy-L-proline; 5'-(N6-L-lysine)-L-topaquinone; 5-hydroxy-L-lysine; 5-methyl-L-arginine; alpha-I-microglobulin-Ig alpha

- complex chromophore; bis-L-cysteinyl bis-L-histidino diiron disulfide; bis-L-cysteinyl-L-N3'-histidino-L-serinyl tetrairon' tetrasulfide; chondroitin sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; D-alanine; D-allo-isoleucine; D-asparagine; dehydroalanine; dehydrotyrosine; dermatan 4-sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; D-glucuronyl-N-glycine; dipyrrolylmethanemethyl-L-cysteine; D-leucine; D-methionine; D-phenylalanine; D-serine; D-tryptophan; glycine amide; glycine oxazolecarboxylic acid; glycine thiazolecarboxylic acid; heme P450-bis-L-cysteine-L-tyrosine; heme-bis-L-cysteine; hemediol-L-aspartyl ester-L-glutamyl ester; hemediol-L-aspartyl ester-L-glutamyl ester-L-methionine sulfonium; heme-L-cysteine; heme-L-histidine; heparan sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; heme P450-bis-L-cysteine-L-lysine; hexakis-L-cysteinyl hexairon hexasulfide; keratan sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-threonine; L-oxoalanine- lactic acid; L phenyllactic acid; l'-(8alpha-FAD)-L-histidine; L-2'.4'.5'-topaquinone; L-3',4'-dihydroxyphenylalanine; L-3'.4'.5'-trihydroxyphenylalanine; L-4'-bromophenylalanine; L-6'-bromotryptophan; L-alanine amide; L-alanyl imidazolinone glycine; L-allysine; L-arginine amide; L-asparagine amide; L-aspartic 4-phosphoric anhydride; L-aspartic acid 1-amide; L-beta-methylthioaspartic acid; L-bromohistidine; L-citrulline; L-cysteine amide; L-cysteine glutathione disulfide; L-cysteine methyl disulfide; L-cysteine methyl ester; L-cysteine oxazolecarboxylic acid; L-cysteine oxazolinecarboxylic acid; L-cysteine persulfide; L-cysteine sulfenic acid; L-cysteine sulfinic acid; L-cysteine thiazolecarboxylic acid; L-cysteinyl homocitryl molybdenum-heptairon-nonasulfide; L-cysteinyl imidazolinone glycine; L-cysteinyl molybdopterin; L-cysteinyl molybdopterin guanine dinucleotide; L-cystine; L-erythro-beta-hydroxyasparagine; L-erythro-beta-hydroxyaspartic acid; L-gamma-carboxyglutamic acid; L-glutamic acid 1-amide; L-glutamic acid 5-methyl ester; L-glutamine amide; L-glutamyl 5-glycerolphosphorylethanolamine; L-histidine amide; L-isoglutamyl-polyglutamic acid; L-isoglutamyl-polyglycine; L-isoleucine amide; L-lanthionine; L-leucine amide; L-lysine amide; L-lysine thiazolecarboxylic acid; L-lysinoalanine; L-methionine amide; L-methionine sulfone; L-phenylalanine thiazolecarboxylic acid; L-phenylalanine amide; L-proline amide; L-selenocysteine; L-selenocysteinyl molybdopterin guanine dinucleotide; L-serine amide; L-serine thiazolecarboxylic acid; L-seryl imidazolinone glycine; L-T-bromophenylalanine; L-T-bromophenylalanine; L-threonine amide; L-thyroxine; L-tryptophan amide; L-tryptophyl quinone; L-tyrosine amide; L-valine amide; meso-

lanthionine; N-(L-glutamyl)-L-tyrosine; N-(L-isoaspartyl)-glycine; N-(L-isoaspartyl)-L-cysteine; N,N,N-trimethyl-L-alanine; N,N-dimethyl-L-proline; N2-acetyl-L-lysine; N2-succinyl-L-tryptophan; N4-(ADP-ribosyl)-L-asparagine; N4-glycosyl-L-asparagine; N4-hydroxymethyl-L-asparagine; N4-methyl-L-asparagine; N5-methyl-L-glutamine; N6- 1 -
5 carboxyethyl-L-lysine; N6-(4-amino hydroxybutyl)-L-lysine; N6-(L-isoglutamyl)-L-lysine; N6-(phospho-5'-adenosine)-L-lysine; N6-(phospho-5'-guanosine)-L-lysine; N6,N6,N6-trimethyl-L-lysine; N6,N6-dimethyl-L-lysine; N6-acetyl-L-lysine; N6-biotinyl-L-lysine; N6-carboxy-L-lysine; N6-formyl-L-lysine; N6-glycyl-L-lysine; N6-lipoyl-L-lysine; N6-methyl-L-lysine; N6-methyl-N6-poly(N-methyl-propylamine)-L-lysine; N6-
10 mureinyl-L-lysine; N6-myristoyl-L-lysine; N6-palmitoyl-L-lysine; N6-pyridoxal phosphate-L-lysine; N6-pyruvic acid 2-iminyl-L-lysine; N6-retinal-L-lysine; N-acetyl-glycine; N-acetyl-L-glutamine; N-acetyl-L-alanine; N-acetyl-L-aspartic acid; N-acetyl-L-cysteine; N-acetyl-L-glutamic acid; N-acetyl-L-isoleucine; N-acetyl-L-methionine; N-acetyl-L-proline; N-acetyl-L-serine; N-acetyl-L-threonine; N-acetyl-L-
15 tyrosine; N-acetyl-L-valine; N-alanyl-glycosylphosphatidylinositoethanolamine; N-asparaginyl-glycosylphosphatidylinositoethanolamine; N-aspartyl-glycosylphosphatidylinositoethanolamine; N-formylglycine; N-formyl-L-methionine; N-glycyl-glycosylphosphatidylinositoethanolamine; N-L-glutamyl-poly-L-glutamic acid; N-methylglycine; N-methyl-L-alanine; N-methyl-L-methionine; N-methyl-L-phenylalanine;
20 N-myristoyl-glycine; N-palmitoyl-L-cysteine; N-pyruvic acid 2-iminyl-L-cysteine; N-pyruvic acid 2-iminyl-L-valine; N-seryl-glycosylphosphatidylinositoethanolamine; N-seryl-glycosylphosphatidylinositoethanolamine; O-(ADP-ribosyl)-L-serine; O-(phospho-5'-adenosine)-L-threonine; O-(phospho-5'-DNA)-L-serine; O-(phospho-5'-DNA)-L-threonine; O-(phospho-5'rRNA)-L-serine; O-(phosphoribosyl dephospho-coenzyme A)-L-serine; O-(sn-l-glycerophosphoryl)-L-serine; O4'-(8alpha-FAD)-L-tyrosine; O4'-(phospho-5'-adenosine)-L-tyrosine; O4'-(phospho-5'-DNA)-L-tyrosine; O4'-(phospho-5'-RNA)-L-tyrosine; O4'-(phospho-5'-uridine)-L-tyrosine; O4-glycosyl-L-hydroxyproline; O4'-glycosyl-L-tyrosine; O4'-sulfo-L-tyrosine; O5-glycosyl-L-hydroxylysine; O-glycosyl-L-serine; O-glycosyl-L-threonine; omega-N-(ADP-ribosyl)-L-arginine; omega-N-omega-N'-
30 dimethyl-L-arginine; omega-N-methyl-L-arginine; omega-N-omega-N-dimethyl-L-arginine; omega-N-phospho-L-arginine; O-octanoyl-L-serine; O-palmitoyl-L-serine; O-palmitoyl-L-threonine; O-phospho-L-serine; O-phospho-L-threonine; O-phosphopantetheine-L-serine; phycoerythrobilin-bis-L-cysteine; phycourobilin-bis-L-

cysteine; pyrroloquinoline quinone; pyruvic acid; S hydroxycinnamyl-L-cysteine; S-(2-aminovinyl) methyl-D-cysteine; S-(2-aminovinyl)-D-cysteine; S-(6-FW-L-cysteine; S-(8alpha-FAD)-L-cysteine; S-(ADP-ribosyl)-L-cysteine; S-(L-isoglutamyl)-L-cysteine; S-12-hydroxyfarnesyl-L-cysteine; S-acetyl-L-cysteine; S-diacylglycerol-L-cysteine; S-5 diphytanylglycerol diether-L-cysteine; S-farnesyl-L-cysteine; S-geranylgeranyl-L-cysteine; S-glycosyl-L-cysteine; S-glycyl-L-cysteine; S-methyl-L-cysteine; S-nitrosyl-L-cysteine; S-palmitoyl-L-cysteine; S-phospho-L-cysteine; S-phytycobiliviolin-L-cysteine; S-phytycyanobilin-L-cysteine; S-phytycoerythrobilin-L-cysteine; S-phytychromobilin-L-cysteine; S-selenyl-L-cysteine; S-sulfo-L-cysteine; tetrakis-L-cysteiny diiron disulfide; 10 tetrakis-L-cysteiny iron; tetrakis-L-cysteiny tetrairon tetrasulfide; trans-2,3-cis 4-dihydroxy-L-proline; tris-L-cysteiny triiron tetrasulfide; tris-L-cysteiny triiron trisulfide; tris-L-cysteiny-L-aspartato tetrairon tetrasulfide; tris-L-cysteiny-L-cysteine persulfido-bis-L-glutamato-L-histidino tetrairon disulfide trioxide; tris-L-cysteiny-L-N3'-histidino tetrairon tetrasulfide; tris-L-cysteiny-L-N1'-histidino tetrairon tetrasulfide; and tris-L-15 cysteiny-L-seriny tetrairon tetrasulfide.

Additional examples of PTMs may be found in web sites such as the Delta Mass database based on Krishna, R. G. and F. Wold (1998). Posttranslational Modifications. Proteins - Analysis and Design. R. H. Angeletti. San Diego, Academic Press. 1: 121-206. ; Methods in Enzymology, 193, J.A. McClosky (ed) (1990), pages 647-660; Methods in 20 Protein Sequence Analysis edited by Kazutomo Imahori and Fumio Sakiyama, Plenum Press, (1993) "Post-translational modifications of proteins" R.G. Krishna and F. Wold pages 167-172; "GlycoSuiteDB: a new curated relational database of glycoprotein/glycan structures and their biological sources" Cooper et al. Nucleic Acids Res. 29; 332-335 (2001) "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta 25 et al. Nucleic Acids Research, 27: 370-372 (1999); and "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. Nucleic Acids Res 27(1):237-239 (1999) see also, WO 02/21139A2, the disclosure of which is incorporated herein by reference in its entirety.

Tumorigenesis is often accompanied by alterations in the post-translational 30 modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known.

One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signalling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell

compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in
5 polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for
10 different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in
15 polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-
20 translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g., p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide
25 may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the
30 polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally

modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule may also be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid
5 sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

It will be appreciated, as is well known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of
10 ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide,
15 including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic
20 processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on
25 the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), e.g., offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546,
30 Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568,

Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY
5 TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

The polypeptides of the present invention can also be conjugated to fluorophores,
10 other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA);
15 common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS,
20 Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available from Pierce, Rockford, IL, USA).

Polypeptides of the present invention, including full length polypeptides, fragments and fusion proteins, can be conjugated, using such cross-linking reagents, to
25 fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to polypeptides of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

Polypeptides of the present invention, including full length polypeptide, fragments and fusion proteins, can also usefully be conjugated using cross-linking agents to carrier
30 proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-OSP antibodies.

Polypeptides of the present invention, including full length polypeptide, fragments and fusion proteins, can also usefully be conjugated to polyethylene glycol (PEG);

PEGylation increases the serum half life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999). PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

Polypeptides of the present invention are also inclusive of analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, this polypeptide is a OSP. In a more preferred embodiment, this polypeptide is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 9-14. Also preferred is an analog polypeptide comprising one or more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally occurring polypeptide. In one embodiment, the analog is structurally similar to a OSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH₂NH--, --CH₂S--, --CH₂-CH₂--, --CH=CH--(cis and trans), --COCH₂--, --CH(OH)CH₂-- and --CH₂SO--. In another embodiment, the analog comprises substitution of one or more amino acids of a OSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g., Koe et al., Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.

Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, *inter alia*, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones,

Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (1993).

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinoyl--(9-fluorenylmethoxycarbonyl)-L-lysine (Fmoc biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The Fmoc and tBOC derivatives of dabcyL-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyL chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyL quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS--Fmoc-L-glutamic acid or the corresponding tBOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated Fmoc synthesis of peptides using (Fmoc)--TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other Fmoc-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, e.g., Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-

2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)- β -alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-
5 hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-
10 methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-
15 (carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical
20 aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified
25 position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

Fusion Proteins

Another aspect of the present invention relates to the fusion of a polypeptide of the present invention to heterologous polypeptides. In a preferred embodiment, the
30 polypeptide of the present invention is a OSP. In a more preferred embodiment, the polypeptide of the present invention that is fused to a heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 9-14, or is a mutein, homologous

polypeptide, analog or derivative thereof. In an even more preferred embodiment, the fusion protein is encoded by a nucleic acid molecule comprising all or part of the nucleic acid sequence of SEQ ID NO: 1-8, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-8.

The fusion proteins of the present invention will include at least one fragment of a polypeptide of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the polypeptide of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of a polypeptide of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and preferably at least 15, 20, or 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particularly useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g.,* Ausubel, Chapter 16, (1992), *supra*. Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins into the periplasmic space or extracellular milieu for prokaryotic hosts or into the culture medium for eukaryotic cells through incorporation of secretion signals and/or leader sequences. For example, a His⁶ tagged protein can be

purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful fusion proteins of the present invention include those that permit use of the polypeptide of the present invention as bait in a yeast two-hybrid system. See Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); Colas *et al.*, *Nature* 380, 548-550 (1996); Norman, T. *et al.*, *Science* 285, 591-595 (1999); Fabbri *et al.*, *Oncogene* 18, 4357-4363 (1999); Xu *et al.*, *Proc Natl Acad Sci U S A* 94, 12473-12478 (1997); Yang, *et al.*, *Nuc. Acids Res.* 23, 1152-1156 (1995); Kolonin *et al.*, *Proc Natl Acad Sci U S A* 95, 14266-14271 (1998); Cohen *et al.*, *Proc Natl Acad Sci U S A* 95, 14272-14277 (1998); Uetz, *et al.* *Nature* 403, 623-627(2000); Ito, *et al.*, *Proc Natl Acad Sci U S A* 98, 4569-4574 (2001). Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded polypeptide on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above.

The polypeptides of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins, β -galactosidase, biotin trpE, protein A, β -lactamase, α -amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast α mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. *See, e.g.*, Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art.

5 Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well known in the art (*e.g.*, a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the OSP.

15

As further described below, the polypeptides of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize polypeptides of the present invention including OSPs and their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly OSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of OSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of OSPs.

20

25

One may determine whether polypeptides of the present invention including OSPs, muteins, homologous proteins or allelic variants or fusion proteins of the present invention are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the polypeptide at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992);

30

combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, (Epicentre

5 Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides or fusion proteins of the present invention is well known and within the skill of one having ordinary skill in the art. *See, e.g.*, Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be
10 effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated polypeptides or fusion proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous and non-proteinaceous material and are well known in the art. Stabilizing
15 agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

Although high levels of purity are preferred when the isolated polypeptide or fusion protein of the present invention are used as therapeutic agents, such as in vaccines and replacement therapy, the isolated polypeptides of the present invention are also useful
20 at lower purity. For example, partially purified polypeptides of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In a preferred embodiment, the purified and substantially purified polypeptides of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

25 The polypeptides or fusion proteins of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent. For example, the peptides of the invention may be stabilized by covalent linkage to albumin. *See*, U.S. Patent No. 5,876,969, the contents of which are hereby incorporated in its entirety.

30 For example, the polypeptides or fusion proteins of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the polypeptides or fusion proteins of the present invention can be used

to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized polypeptide or fusion protein of the present invention.

As another example, the polypeptides or fusion proteins of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides and fusion proteins of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the polypeptide or fusion protein of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound polypeptide or fusion protein to indicate biologic interaction there between. The polypeptides or fusion proteins of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the polypeptide or fusion protein of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound polypeptide or fusion protein to indicate biological interaction there between.

Alternative Transcripts

In another aspect, the present invention provides splice variants of genes and proteins encoded thereby. The identification of a novel splice variant which encodes an amino acid sequence with a novel region can be targeted for the generation of reagents for use in detection and/or treatment of cancer. The novel amino acid sequence may lead to a unique protein structure, protein subcellular localization, biochemical processing or function of the splice variant. This information can be used to directly or indirectly facilitate the generation of additional or novel therapeutics or diagnostics. The nucleotide sequence in this novel splice variant can be used as a nucleic acid probe for the diagnosis and/or treatment of cancer.

Specifically, the newly identified sequences may enable the production of new antibodies or compounds directed against the novel region for use as a therapeutic or

diagnostic. Alternatively, the newly identified sequences may alter the biochemical or biological properties of the encoded protein in such a way as to enable the generation of improved or different therapeutics targeting this protein.

Antibodies

5 In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention. In a preferred embodiment, the antibodies are specific for a polypeptide that is a OSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that
10 comprises SEQ ID NO: 9-14, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on
15 the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a OSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or vis versa. In addition, alternative splice forms of a OSP may be indicative of cancer. Differential degradation of the C or N-terminus of a
20 OSP may also be a marker or target for anticancer therapy. For example, an OSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is well known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational
25 relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-OSP polypeptides by at least two-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the
30 antibody of the present invention is sufficiently specific when it can be used to determine the presence of the polypeptide of the present invention in samples derived from human ovarian.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7} M, 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and up to 1×10^{-13} M proving especially useful.

The antibodies of the present invention can be naturally occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In such case, antibodies to the polypeptides of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the polypeptide of the present invention. Such antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patent Nos. 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention are also usefully obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster), lagomorphs (typically rabbits), and also larger mammals, such as sheep, goats, cows, and horses; or egg laying birds or reptiles such as chickens or alligators. In such cases, as with the transgenic human-antibody-

producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the polypeptide of the present invention. One form of avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000.

5 As discussed above, virtually all fragments of 8 or more contiguous amino acids of a polypeptide of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

10 Immunogenicity can also be conferred by fusion of the polypeptide of the present invention to other moieties. For example, polypeptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

15 Protocols for immunizing non-human mammals or avian species are well-established in the art. See Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996). Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin.* 25 *Immunol.* 2: 317-327 (1990).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the polypeptides of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the polypeptides of the present invention. Antibodies from avian species may have particular advantage in detection of the polypeptides of the present invention, in human serum or tissues (Vikinge *et al.*, *Biosens. Bioelectron.* 13: 1257-1262 (1998). Following immunization, the antibodies of the present invention can be obtained using any

Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell. Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention. For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. See, e.g., Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, *J. Biotechnol.* 76(2-3):1 57-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); and Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998).

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. See, e.g., Li *et al.*, *Protein Expr. Purif.* 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997); and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992).

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995).

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. See, e.g. Pollock *et al.*, *J. Immunol Methods.* 231: 147-57 (1999); Young *et al.*, *Res. Immunol.* 149: 609-10 (1998); and Limonta *et al.*, *Immunotechnology* 1: 107-13 (1995).

Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells. Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998) review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies. Antibodies of the present invention can also be prepared by cell

free translation, as further described in Merk *et al.*, *J. Biochem.* (Tokyo) 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57 (1999).

5 The invention further provides antibody fragments that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid
10 molecules of the present invention. Among such useful fragments are Fab, Fab', Fv, F(ab)'₂, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

 The present invention also relates to antibody derivatives that bind specifically to one or more of the polypeptides of the present invention, to one or more of the
15 polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention.

 Among such useful derivatives are chimeric, primatized, and humanized
20 antibodies; such derivatives are less immunogenic in human beings, and thus are more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful method is PEGylation to increase the serum half life of the antibodies.

 Chimeric antibodies typically include heavy and/or light chain variable regions
25 (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*, Morrison *et al.*, *Proc. Natl. Acad. Sci USA* 81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985); and U.S. Patent No. 5,807,715 the disclosure of which is incorporated herein by reference in its entirety.
30 Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); and U.S. Patent Nos.

6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties. Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies),
5 single-chain diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. Accordingly, the present invention includes any recombinant vector containing the coding sequences, or part
10 thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be
15 transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives
20 thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the
25 polypeptides encoded by the isolated nucleic acid molecules of the present invention. The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label can usefully be an enzyme that catalyzes production and local deposition of a detectable product. Enzymes typically
30 conjugated to antibodies to permit their immunohistochemical visualization are well known, and include alkaline phosphatase, β -galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG);

o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopyranoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BlueGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide (H₂O₂), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. See, e.g., Thorpe *et al.*, *Methods Enzymol.* 133: 331-53 (1986); Kricka *et al.*, *J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al.*, *J. Biolumin. Chemilumin.* 10(6): 353-9 (1995). Kits for such enhanced chemiluminescent detection (ECL) are available commercially. The antibodies can also be labeled using colloidal gold.

As another example, when the antibodies of the present invention are used, e.g., for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores. There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention. For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY

564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention. For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P , ^{32}P , ^{35}S , ^3H , and ^{125}I . As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th , ^{227}Ac , ^{225}Ac , ^{223}Ra , ^{213}Bi , ^{212}Pb , ^{212}Bi , ^{211}At , ^{203}Pb , ^{194}Os , ^{188}Re , ^{186}Re , ^{153}Sm , ^{149}Tb , ^{131}I , ^{125}I , ^{111}In , ^{105}Rh , $^{99\text{m}}\text{Tc}$, ^{97}Ru , ^{90}Y , ^{90}Sr , ^{88}Y , ^{72}Se , ^{67}Cu , or ^{47}Sc .

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the application as for which they were mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the polypeptides of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, or ricin. See Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998).

The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic

acid molecules of the present invention, attached to a substrate. Substrates can be porous or nonporous, planar or nonplanar. For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography. For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microsphere can then be used for isolation of cells that express or display the polypeptides of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the OSPs of the present invention or to polypeptides encoded by the OSNAs of the invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding a OSP. In a preferred embodiment, the OSP comprises an amino acid sequence selected from SEQ ID NO: 9-14, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise a OSNA of the invention, preferably a OSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-8, or a part,

substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human OSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well known in the art. *See, e.g., Hogan et al., Manipulating the Mouse Embryo: A Laboratory Manual*, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, *Mouse Genetics and Transgenics: A Practical Approach*, Oxford University Press (2000); and Pinkert, *Transgenic Animal Technology: A Laboratory Handbook*, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g., Paterson et al., Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11: 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and U.S. Patent No. 4,873,191, herein incorporated by reference in its entirety); retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g., Van der Putten et al., Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g., Thompson et al., Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g., Lo, 1983, Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g., Ulmer et al., Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g., Lavitrano et al., Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g., Campbell et al., Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e., a nucleic acid molecule of the invention*) in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i.e. e., mosaic animals or chimeric animals.*

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, *e. g., head-to-head tandems or head-to-tail tandems.* The transgene may also be selectively introduced into and activated in a particular cell type by following,

e.g., the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232- 6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant
5 gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained
10 from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or
15 crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous
20 transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

25 Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

30 Methods for creating a transgenic animal with a disruption of a targeted gene are also well known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences,

into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g., Gu et al., Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will
5 depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g., Smithies et al., Nature* 317: 230-234 (1985); Thomas *et al., Cell* 51: 503-512 (1987); Thompson *et al., Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the
10 endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via
15 targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g., Thomas, supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly
20 administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g., knockouts*) are administered to a
25 patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g., lymphocytes*), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt
30 the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g., by transduction* (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures,

including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

5 The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

10 Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.*, U.S. Patent Nos. 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

15 When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

20 Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

25 A further aspect of the invention is a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 9-14 and SEQ ID NO: 1-8 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading
30 and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the

comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis,

sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a
5 nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence
10 comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the
15 steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence. In addition, the invention includes a method of using patterns of expression associated with either the nucleic acids or proteins in a computer-based method to
20 diagnose disease.

Diagnostic Methods for ovarian Cancer

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by comparing expression of a OSNA or a OSP in a human patient that has or may have
25 ovarian cancer, or who is at risk of developing ovarian cancer, with the expression of a OSNA or a OSP in a normal human control. For purposes of the present invention, "expression of a OSNA" or "OSNA expression" means the quantity of OSNA mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient.
30 Similarly, the term "expression of a OSP" or "OSP expression" means the amount of OSP that can be measured by any method known in the art or the level of translation of a OSNA that can be measured by any method known in the art.

The present invention provides methods for diagnosing ovarian cancer in a patient, by analyzing for changes in levels of OSNA or OSP in cells, tissues, organs or bodily fluids compared with levels of OSNA or OSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of a OSNA or OSP in the patient versus the normal human control is associated with the presence of ovarian cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing ovarian cancer in a patient by analyzing changes in the structure of the mRNA of a OSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing ovarian cancer in a patient by analyzing changes in a OSP compared to a OSP from a normal patient. These changes include, *e.g.*, alterations, including post translational modifications such as glycosylation and/or phosphorylation of the OSP or changes in the subcellular OSP localization.

For purposes of the present invention, diagnosing means that OSNA or OSP levels are used to determine the presence or absence of disease in a patient. As will be understood by those of skill in the art, measurement of other diagnostic parameters may be required for definitive diagnosis or determination of the appropriate treatment for the disease. The determination may be made by a clinician, a doctor, a testing laboratory, or a patient using an over the counter test. The patient may have symptoms of disease or may be asymptomatic. In addition, the OSNA or OSP levels of the present invention may be used as screening marker to determine whether further tests or biopsies are warranted. In addition, the OSNA or OSP levels may be used to determine the vulnerability or susceptibility to disease.

In a preferred embodiment, the expression of a OSNA is measured by determining the amount of a mRNA that encodes an amino acid sequence selected from SEQ ID NO: 9-14, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the OSNA expression that is measured is the level of expression of a OSNA mRNA selected from SEQ ID NO: 1-8, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acid molecules. OSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by Northern blot, quantitative or qualitative

reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.,* Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. OSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of a OSG of interest or
5 doing nuclear run-off assays. Alterations in mRNA structure, *e.g.,* aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, OSNA expression may be compared to a known control, such as normal ovarian nucleic acid, to detect a change in expression.

In another preferred embodiment, the expression of a OSP is measured by
10 determining the level of a OSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 9-14, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression
15 of a OSNA or OSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of ovarian cancer. The expression level of a OSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the OSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry,
20 immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See, e.g.,* Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the OSP structure may be determined by any method known in the art, including, *e.g.,* using antibodies that specifically recognize phosphoserine,
25 phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein.
Id.

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to a OSP is prepared if one is not already available. In a preferred
30 embodiment, the antibody is a monoclonal antibody. The anti-OSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the OSP will bind to the anti-OSP antibody.

The sample is removed, the solid support is washed to remove unbound material, and an anti-OSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the OSP to the labeled antibody will occur. After binding, the unbound
5 labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of an OSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

10 Other methods to measure OSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-OSP antibody is attached to a solid support and an allocated amount of a labeled OSP and a sample of interest are incubated with the solid support. The amount of labeled OSP attached to the solid support can be correlated to the quantity of a OSP in the sample.

15 Of the proteomic approaches, 2D PAGE is a well known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is
20 perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

25 Expression levels of a OSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a
30 specific mRNA population in a complex mixture of thousands of other mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more OSNAs of interest. In this approach, all or a portion of one or more OSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*,
5 total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without
10 limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy
15 material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. As used herein "blood" includes whole blood, plasma, serum, circulating epithelial cells, constituents, or any derivative of blood.

In addition to detection in bodily fluids, the proteins and nucleic acids of the invention are suitable to detection by cell capture technology. Whole cells may be
20 captured by a variety methods for example magnetic separation, U.S. Patent. Nos. 5,200,084; 5,186,827; 5,108,933; 4,925,788, the disclosures of which are incorporated herein by reference in their entirety. Epithelial cells may be captured using such products as Dynabeads® or CELlection™ (DynaL Biotech, Oslo, Norway). Alternatively, fractions of blood may be captured, *e.g.*, the buffy coat fraction (50mm cells isolated from
25 5ml of blood) containing epithelial cells. In addition, cancer cells may be captured using the techniques described in WO 00/47998, the disclosure of which is incorporated herein by reference in its entirety. Once the cells are captured or concentrated, the proteins or nucleic acids are detected by the means described in the subject application. Alternatively, nucleic acids may be captured directly from blood samples, see U.S. Patent Nos.
30 6,156,504, 5,501,963; or WO 01/42504, the disclosures of which are incorporated herein by reference in their entirety.

In a preferred embodiment, the specimen tested for expression of OSNA or OSP includes without limitation ovarian tissue, ovarian cells grown in cell culture, blood,

serum, lymph node tissue, and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary ovarian cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, lungs, colon, and adrenal glands. In general, the tissues may be sampled by biopsy, including, without
5 limitation, needle biopsy, e.g., transthoracic needle aspiration, cervical mediastinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the
10 expression level of a OSNA or OSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other OSNA or OSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer
15 marker in addition to a particular OSNA or OSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

20 In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more OSNA and/or OSP in a sample from a patient suspected of having ovarian cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of a OSNA and/or OSP and then ascertaining whether the patient has ovarian
25 cancer from the expression level of the OSNA or OSP. In general, if high expression relative to a control of a OSNA or OSP is indicative of ovarian cancer, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least one and a half times higher, and more preferably are at least two times higher, still more preferably five times higher, even more preferably at least ten times higher, than in preferably the
30 same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a OSNA or OSP is indicative of ovarian cancer, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at

least one and a half times lower, and more preferably are at least two times lower, still more preferably five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether ovarian cancer has metastasized in a patient. One may identify whether the ovarian cancer has metastasized by measuring the expression levels and/or structural alterations of one or more OSNAs and/or OSPs in a variety of tissues. The presence of a OSNA or OSP in a certain tissue at levels higher than that of corresponding noncancerous tissue (e.g., the same tissue from another individual) is indicative of metastasis if high level expression of a OSNA or OSP is associated with ovarian cancer. Similarly, the presence of a OSNA or OSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of a OSNA or OSP is associated with ovarian cancer. Further, the presence of a structurally altered OSNA or OSP that is associated with ovarian cancer is also indicative of metastasis.

In general, if high expression relative to a control of a OSNA or OSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the OSNA or OSP is at least one and a half times higher, and more preferably are at least two times higher, still more preferably five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a OSNA or OSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the OSNA or OSP is at least one and a half times lower, and more preferably are at least two times lower, still more preferably five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

Staging

The invention also provides a method of staging ovarian cancer in a human patient. The method comprises identifying a human patient having ovarian cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more OSNAs or OSPs. First, one or more tumors from a

variety of patients are staged according to procedures well known in the art, and the expression levels of one or more OSNAs or OSPs is determined for each stage to obtain a standard expression level for each OSNA and OSP. Then, the OSNA or OSP expression levels of the OSNA or OSP are determined in a biological sample from a patient whose stage of cancer is not known. The OSNA or OSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the OSNAs and OSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of a OSNA or OSP to determine the stage of a ovarian cancer.

10 *Monitoring*

Further provided is a method of monitoring ovarian cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the ovarian cancer. The monitoring may determine if there has been a reoccurrence and, if so, determine its nature. The method comprises identifying a human patient that one wants to monitor for ovarian cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more OSNAs or OSPs, and comparing the OSNA or OSP levels over time to those OSNA or OSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in a OSNA or OSP that are associated with ovarian cancer.

If increased expression of a OSNA or OSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of a OSNA or OSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of a OSNA or OSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting a decrease in the expression level of a OSNA or OSP indicates that

the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of OSNAs or OSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples. Monitoring a patient for onset of ovarian cancer metastasis is periodic and preferably is done on a
5 quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of a OSNA and/or OSP. The present invention provides a method in which a test sample is obtained from a human patient and one or
10 more OSNAs and/or OSPs are detected. The presence of higher (or lower) OSNA or OSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly ovarian cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more OSNAs and/or OSPs of the invention can also be monitored by analyzing levels of expression of the OSNAs
15 and/or OSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

Detection of Genetic Lesions or Mutations

20 The methods of the present invention can also be used to detect genetic lesions or mutations in a OSG, thereby determining if a human with the genetic lesion is susceptible to developing ovarian cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing ovarian cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of
25 one or more nucleotides from the OSGs of this invention, a chromosomal rearrangement of a OSG, an aberrant modification of a OSG (such as of the methylation pattern of the genomic DNA), or allelic loss of a OSG. Methods to detect such lesions in the OSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

30 Methods of Detecting Noncancerous ovarian Diseases

The present invention also provides methods for determining the expression levels and/or structural alterations of one or more OSNAs and/or OSPs in a sample from a

patient suspected of having or known to have a noncancerous ovarian disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of a OSNA and/or OSP, comparing the expression level or structural alteration of the OSNA or OSP to a normal ovarian control, and then ascertaining whether the patient has a noncancerous ovarian disease. In general, if high expression relative to a control of a OSNA or OSP is indicative of a particular noncancerous ovarian disease, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a OSNA or OSP is indicative of a noncancerous ovarian disease, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether a OSNA and/or OSP is associated with a particular noncancerous ovarian disease by obtaining ovarian tissue from a patient having a noncancerous ovarian disease of interest and determining which OSNAs and/or OSPs are expressed in the tissue at either a higher or a lower level than in normal ovarian tissue. In another embodiment, one may determine whether a OSNA or OSP exhibits structural alterations in a particular noncancerous ovarian disease state by obtaining ovarian tissue from a patient having a noncancerous ovarian disease of interest and determining the structural alterations in one or more OSNAs and/or OSPs relative to normal ovarian tissue.

Methods for Identifying ovarian Tissue

In another aspect, the invention provides methods for identifying ovarian tissue. These methods are particularly useful in, *e.g.*, forensic science, ovarian cell differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is ovarian tissue or has ovarian tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising ovarian tissue or having ovarian

tissue-like characteristics, determining whether the sample expresses one or more OSNAs and/or OSPs, and, if the sample expresses one or more OSNAs and/or OSPs, concluding that the sample comprises ovarian tissue. In a preferred embodiment, the OSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 9-14, or a
5 homolog, allelic variant or fragment thereof. In a more preferred embodiment, the OSNA has a nucleotide sequence selected from SEQ ID NO: 1-8, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses a OSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In
10 another preferred embodiment, the method can be practiced by determining whether a OSP is expressed. Determining whether a sample expresses a OSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the OSP has an amino acid sequence selected from SEQ ID NO: 9-14, or a homolog, allelic variant or fragment thereof. In another preferred
15 embodiment, the expression of at least two OSNAs and/or OSPs is determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five OSNAs and/or OSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is ovarian tissue. This is particularly useful in forensic science, in which small,
20 damaged pieces of tissues that are not identifiable by microscopic or other means are recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into ovarian tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new ovarian tissue by tissue engineering. These agents
25 include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

Methods for Producing and Modifying ovarian Tissue

30 In another aspect, the invention provides methods for producing engineered ovarian tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing a OSNA or a OSG into the cells, and growing the cells under conditions

in which they exhibit one or more properties of ovarian tissue cells. In a preferred embodiment, the cells are pluripotent. As is well known in the art, normal ovarian tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered ovarian tissue or cells comprises one of these cell types. In another
5 embodiment, the engineered ovarian tissue or cells comprises more than one ovarian cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the ovarian cell tissue. Methods for manipulating culture conditions are well known in the art.

Nucleic acid molecules encoding one or more OSPs are introduced into cells,
10 preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode OSPs having amino acid sequences selected from SEQ ID NO: 9-14, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1-8, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly
15 preferred embodiment, a OSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well known in the art and are described in detail, *supra*.

Artificial ovarian tissue may be used to treat patients who have lost some or all of their ovarian function.

20 Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, fusion proteins, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, or inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises a OSNA or part thereof.
25 In a more preferred embodiment, the OSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-8, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises a OSP or fragment thereof. In a more preferred embodiment, the pharmaceutical composition comprises a
30 OSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 9-14, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred

embodiment, the pharmaceutical composition comprises an anti-OSP antibody, preferably an antibody that specifically binds to a OSP having an amino acid that is selected from the group consisting of SEQ ID NO: 9-14, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art that is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000) and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, cornstarch, sodium starch glycolate, and alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline,

0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form
5 of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil,
10 triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

15 Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

20 Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of
25 other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically. For topical use the compounds of the present invention can also be prepared in
30 suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the

active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from
5 about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be
10 administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such
15 as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of
delivery.

20 The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

25 After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

30 A "therapeutically effective dose" refers to that amount of active ingredient, for example OSP polypeptide, fusion protein, or fragments thereof, antibodies specific for OSP, agonists, antagonists or inhibitors of OSP, which ameliorates the signs or symptoms

of the disease or prevent progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed
5 by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one
10 or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of
15 circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

20 The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting
25 pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody
30 agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

10 Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention, *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of ovarian function. As used herein, "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

Gene Therapy and Vaccines

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV), for the purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of "naked" nucleic acid vaccination, as further described in U.S. Patent Nos. 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g.*, Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid molecule of the present invention is administered. The nucleic acid molecule can be

delivered in a vector that drives expression of a OSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of a OSP are administered, for example, to complement a deficiency in the native OSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g., Cid-Arregui, supra.* In a preferred embodiment, the nucleic acid molecule encodes a OSP having the amino acid sequence of SEQ ID NO: 9-14, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express a OSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in OSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode a OSP having the amino acid sequence of SEQ ID NO: 9-14, or a fragment, fusion protein, allelic variant or homolog thereof.

Antisense Administration

Antisense nucleic acid compositions, or vectors that drive expression of a OSG antisense nucleic acid, are administered to downregulate transcription and/or translation of a OSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of a OSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.,* between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to OSG transcripts, are also useful in therapy. *See, e.g., Phylactou, Adv. Drug Deliv. Rev. 44(2-3): 97-108 (2000); Phylactou et al., Hum. Mol. Genet. 7(10): 1649-53 (1998); Rossi, Ciba Found. Symp. 209: 195-204 (1997); and Sigurdsson et al., Trends Biotechnol. 13(8): 286-9 (1995).*

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the OSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g., Intody et al., Nucleic*

Acids Res. 28(21): 4283-90 (2000); and McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000). Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

5 In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding a OSP, preferably a OSP comprising an amino acid sequence of SEQ ID NO: 9-14, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-8, or a part, allelic variant, substantially similar or
10 hybridizing nucleic acid thereof.

Polypeptide Administration

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a OSP, a fusion protein, fragment, analog or derivative thereof is administered to a subject with a
15 clinically-significant OSP defect.

Protein compositions are administered, for example, to complement a deficiency in native OSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to OSP. The immune response can be used to modulate activity of OSP or, depending on the immunogen, to immunize against
20 aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate OSP.

In a preferred embodiment, the polypeptide administered is a OSP comprising an amino acid sequence of SEQ ID NO: 9-14, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded
25 by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-8, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a
30 therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well known, antibody compositions are administered, for example, to antagonize activity

of OSP, or to target therapeutic agents to sites of OSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to a OSP comprising an amino acid sequence of SEQ ID NO: 9-14, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antibody specifically binds to a
5 OSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-8, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to a OSP or have a modulatory effect on the expression or activity of a OSP. Modulators which decrease the expression or activity of OSP (antagonists) are believed to
10 be useful in treating ovarian cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of a OSP can also be designed, synthesized and tested for use in the imaging and treatment of ovarian cancer. Further, libraries of molecules can be screened for potential anticancer agents by
15 assessing the ability of the molecule to bind to the OSPs identified herein. Molecules identified in the library as being capable of binding to a OSP are key candidates for further evaluation for use in the treatment of ovarian cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of a OSP in cells.

In another embodiment of the therapeutic methods of the present invention, a
20 pharmaceutical composition comprising a non-antibody antagonist of OSP is administered. Antagonists of OSP can be produced using methods generally known in the art. In particular, purified OSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of a OSP.

25 In other embodiments a pharmaceutical composition comprising an agonist of a OSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a OSP comprising an amino acid sequence of SEQ
30 ID NO: 9-14, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a OSP encoded by a nucleic acid molecule having a

nucleotide sequence of SEQ ID NO: 1-8, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting ovarian Tissue

5 The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the ovarian or to specific cells in the ovarian. In a preferred embodiment, an anti-OSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if ovarian tissue needs to be
10 selectively destroyed. This would be useful for targeting and killing ovarian cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting ovarian cell function.

 In another embodiment, an anti-OSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be
15 useful for determining and monitoring ovarian function, identifying ovarian cancer tumors, and identifying noncancerous ovarian diseases.

EXAMPLES

Example 1: Gene Expression analysis

Example 1: Alternative Splice Variants

20 We identified gene transcripts associated with cancer disease, development, or progression using cloning experiments, the Gencarta™ tools software (Compugen Ltd., Tel Aviv, Israel), and a variety of public and proprietary databases. These splice variants are either novel sequences which differ from a previously defined sequence or new uses of known sequences. In general the previously defined sequence for a family is annotated as
25 DEX0439_XXX.nt.1 and the novel variants are annotated as DEX0439_XXX.nt.2, DEX0439_XXX.nt.3, etc. The novel variant DNA sequences encode novel proteins which differ from a previously defined protein sequence. In relation to the nucleotide sequence naming convention, the previously defined amino acid sequence is annotated DEX0439_XXX.aa.1 and the novel variants annotated as DEX0439_XXX.aa.2, etc.

The mapping of the nucleic acid ("NT") SEQ ID NO; DEX ID; DEX lead name, chromosomal location (if known); open reading frame (ORF) location; amino acid ("AA") SEQ ID NO and AA DEX ID are shown in the table below.

SEQ ID NO	DEX NT SEQ ID	Lead Name	Chromo Map	ORF Loc	SEQ ID NO	DEX AA SEQ ID
1	DEX0439_001.nt.1	Ovr107	19q13.42	103-1890	9	DEX0439_001.aa.1
2	DEX0439_001.nt.2		19q13.42	870-1370	10	DEX0439_001.aa.2
2	DEX0439_001.nt.2		19q13.42	1-1224	11	DEX0439_001.aa.3
3	DEX0439_001.nt.4	Ovr107v4	19q13.42	13-1422	10	DEX0439_001.aa.2
3	DEX0439_001.nt.4	Ovr107v4	19q13.42	1-1206	11	DEX0439_001.aa.3
4	DEX0439_001.nt.6		19q13.42	13-1422	10	DEX0439_001.aa.2
4	DEX0439_001.nt.6		19q13.42	1-1206	11	DEX0439_001.aa.3
5	DEX0439_001.nt.8		19q13.42	13-1422	10	DEX0439_001.aa.2
5	DEX0439_001.nt.8		19q13.42	1-1206	11	DEX0439_001.aa.3
6	DEX0439_001.nt.10		19q13.42	13-1422	10	DEX0439_001.aa.2
6	DEX0439_001.nt.10		19q13.42	1-1206	11	DEX0439_001.aa.3
7	DEX0439_002.nt.1	Ovr110	20p11.21	71-916	12	DEX0439_002.aa.1
8	DEX0439_002.nt.2	Ovr110v1	20p11.21	73-249	13	DEX0439_002.aa.2
8	DEX0439_002.nt.2	Ovr110v1	20p11.21	253-1008	14	DEX0439_002.aa.3

a. EST Support

- 5 The alternative splice variants were predicted by computational analysis of Expressed Sequence Tags (ESTs) derived from public and proprietary cDNA libraries and genomic information. The tissue distribution and number of ESTs which support each alternative splice variant sequence is provided in the transcript sections below.

b. SAGE Support

- 10 Serial Analysis of Gene Expression (SAGE) tag data analysis was also performed on the novel splice variants. Gencarta™ tools (Compugen Ltd., Tel Aviv, Israel) report SAGE tag data for individual transcripts when available. SAGE data reported includes the SAGE tag sequence for the novel transcripts, expression level (as a ratio) of the SAGE tag in tissue samples, the source or tissue, state or disease condition of the tissue, tissue

sample type, and a description of the tissue samples. SAGE tag data analysis results are disclosed and discussed in each transcript section below.

c. Sequence Alignment Support

Alignments of previously identified and novel splice variant sequences were performed to confirm unique portions of splice variant nucleic acid and amino acid sequences. The alignments were done using the Needle program in the European Molecular Biology Open Software Suite (EMBOSS) version 2.2.0 available at www.emboss.org from EMBnet (<http://www.embnet.org>). Default settings were used unless otherwise noted. The Needle program in EMBOSS implements the Needleman-Wunsch algorithm. Needleman, S. B., Wunsch, C. D., *J. Mol. Biol.* 48:443-453 (1970).

It is well known to those skilled in the art that implication of alignment algorithms by various programs may result in minor changes in the generated output. These changes include but are not limited to: alignment scores (percent identity, similarity, and gap), display of nonaligned flanking sequence regions, and number assignment to residues. These minor changes in the output of an alignment do not alter the physical characteristics of the sequences or the differences between the sequences, e.g. regions of homology, insertions, or deletions. Descriptions of alignments are provided in each splice variant section below.

d. RT-PCR Analysis

To detect the presence and tissue distribution of a particular splice variant Reverse Transcription-Polymerase Chain Reaction (RT-PCR) was performed using cDNA generated from a panel of tissue RNAs. See, e.g., Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and; Kawasaki ES *et al.*, *PNAS* 85(15):5698 (1988). Total RNA was extracted from a variety of tissues and first strand cDNA was prepared with reverse transcriptase (RT). Each panel includes 23 cDNAs from five cancer types (lung, ovary, breast, colon, and prostate) and normal samples of testis, placenta and fetal brain. Each cancer set is composed of three cancer cDNAs from different donors and one normal pooled sample. Using a standard enzyme kit from BD Bioscience Clontech (Mountain View, CA), the target transcript was detected with sequence-specific primers designed to only amplify the particular splice variant. The PCR reaction was run on the GeneAmp PCR system 9700 (Applied Biosystem, Foster City, CA) thermocycler under optimal conditions. One of ordinary skill can design

appropriate primers and determine optimal conditions. The amplified product was resolved on an agarose gel to detect a band of equivalent size to the predicted RT-PCR product. A band indicated the presence of the splice variant in a sample. The relation of the amplified product to the splice variant was subsequently confirmed by DNA

5 sequencing.

RT-PCR results are disclosed and discussed in each transcript section below. The RT-PCR experiments confirm the physical existence of the alternative splice variants, SEQ ID NO: 1-8 in a biological sample.

e. Relative Quantitation of Gene Expression Analysis

10 Real-Time quantitative PCR with fluorescent Taqman[®] probes is a quantitation detection system utilizing the 5'-3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman[®]) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by
15 the laser detector of the Model 7900 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this endogenous control.
20 To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the comparative method (User Bulletin #2: ABI PRISM 7900 Sequence Detection System).

The tissue distribution and the level of the target gene are evaluated for every
25 sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman[®] probes specific to each target gene. The results are analyzed using the ABI PRISM 7900 Sequence Detector. The absolute numbers are
30 relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the OSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to the calibrator. Normal RNA samples are commercially available pools, originated by pooling samples of a particular tissue from
5 different individuals.

The relative levels of expression of the OSNA in pairs of matched samples may also be determined. A matched pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. All the values are compared to the calibrator.

10 In the analysis of matching samples, the OSNAs show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples. Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer state (e.g. higher
15 levels of mRNA expression in the cancer sample compared to the normal adjacent).

Real-Time quantitative PCR results are disclosed and discussed in each transcript section below. Altogether, the high level of tissue specificity, plus the mRNA overexpression in matched samples tested are indicative of SEQ ID NO: 1-8 being a diagnostic marker and/or a therapeutic target for cancer.

20 *f. Secretion Assay Analysis*

To determine if a protein encoded by a splice variant is secreted from cells a secretion assay is preformed. A pcDNA3.1 clone containing the gene transcript which encodes the variant protein is transfected into 293T cells using the Superfect transfection reagent (Qiagen, Valencia CA). Transfected cells are incubated for 28 hours before the
25 media is collected and immediately spun down to remove any detached cells. The adherent cells are solubilized with lysis buffer (1% NP40, 10mM sodium phosphate pH7.0, and 0.15M NaCl). The lysed cells are collected and spun down and the supernatant extracted as cell lysate. Western immunoblot is carried out in the following manner: 15µl of the cell lysate and media are run on 4-12% NuPage Bis-Tris gel
30 (Invitrogen, Carlsbad CA), and blotted onto a PVDF membrane (Invitrogen, Carlsbad CA). The blot is incubated with a polyclonal primary antibody which binds to the variant protein (Imgenex, San Diego CA) and polyclonal goat anti-rabbit-peroxidase secondary

antibody (Sigma-Aldrich, St. Louis MO). The blot is developed with the ECL Plus chemiluminescent detection reagent (Amersham BioSciences, Piscataway NJ).

5 Secretion assay results are disclosed and discussed in each transcript section below. Secretion assay results are indicative of SEQ ID NO: 9-14 being a diagnostic marker and/or therapeutic target for cancer.

DEX0439_001.nt.1 (Ovr107); Splice Variants DEX0439_001.nt.2, DEX0439_00X.nt.4 (Ovr107v4), DEX0439_001.nt.6, DEX0439_001.nt.8, DEX0439_001.nt.10

10 Ovr107 was previously identified wholly or in part as Human Ovr107 ovarian cancer marker DNA in WO 01/37864-A1; Human bone remodelling gene #51 in US 6,426,186-B1; Human cDNA SEQ ID NO 71 in WO 02/18424-A2 and Human full-length cDNA, SEQ ID NO: 3266 in EP1130094-A2 which are herein incorporated by reference.

15 Ovr107 is related to Homo sapiens EPS8-like 1 (EPS8L1), transcript variant 2, mRNA. This gene encodes a protein that is related to epidermal growth factor receptor pathway substrate 8 (EPS8), a substrate for the epidermal growth factor receptor. The function of this protein is unknown. Several alternatively spliced transcript variants encoding different isoforms exist. This variant (2) differs in the 5' UTR and contains an alternate in-frame segment in the 5' coding region compared to variant 1. It encodes a shorter isoform (b) with a different N-terminus compared to isoform a. (RefSeq ID: NM_017729.2). Tocchetti, A. *et al.*, *Genomics* 81:234-244 (2003); Wu, K. *et al.*, *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 16:325-327 (1999).

20 Splice variants have been identified for Ovr107 using the methods described above. They are DEX0439_001.nt.2, DEX0439_00X.nt.4 (Ovr107v4), DEX0439_001.nt.6, DEX0439_001.nt.8, DEX0439_001.nt.10. These transcripts arise from alternative splicing events in the same genomic region as Ovr107 and contain exons encoding amino acid sequences. These amino acid sequences provide proteins to be targeted for the generation of reagents that can be used in the detection and/or treatment of cancer. The nucleotide sequences in these exons can be used as a nucleic acid probe for the diagnosis and/or treatment of cancer.

30 DEX0439_001.nt.2 Splice Variant

DEX0439_001.nt.2 is a protein encoding sequence supported by 2 ESTs from cervical dysplasia and fetal liver tissues. This splice variant contains exons that

distinguish it from DEX0439_001.nt.1 (Ovr107). An alignment of the DNA sequences for DEX0439_001.nt.1 (Ovr107) and DEX0439_001.nt.2 is provided in Figure 1.

DEX0439_001.nt.2 encodes the amino acid sequence DEX0439_001.aa.2 which comprises insertions and deletions that distinguish it from DEX0439_001.aa.1 (Ovr107).

5 An alignment of the protein sequences for DEX0439_001.aa.1 (Ovr107) and DEX0439_001.aa.2 is provided in Figure 2.

DEX0439_001.nt.2 encodes the amino acid sequence DEX0439_001.aa.3 which comprises insertions and deletions that distinguish it from DEX0439_001.aa.1 (Ovr107).

10 An alignment of the protein sequences for DEX0439_001.aa.1 (Ovr107) and DEX0439_001.aa.3 is provided in Figure 3.

DEX0439_001.nt.2 RT-PCR Analysis

RT-PCR experiments were performed according to the methods described above in subsection d. Primers were designed unique to DEX0439_001.nt.2. Results of RT-PCR confirmed the presence of the transcript in a biological sample.

15 After subcloning, all positively screened clones were sequence verified. DNA sequence verification confirmed DEX0439_001.nt.2 is distinct from Ovr107.

DEX0439_00X.nt.4 (Ovr107v4) Splice Variant

Ovr107v4 is a protein encoding sequence supported by 9 ESTs from papillary carcinoma clones and ovarian tumor tissue. This splice variant contains exons that distinguish it from Ovr107. An alignment of the DNA sequences for DEX0439_001.nt.1 (Ovr107) and DEX0439_00X.nt.4 (Ovr107v4) is provided in Figure 4.

20 DEX0439_00X.nt.4 (Ovr107v4) encodes the amino acid sequence DEX0439_001.aa.2 which comprises insertions and deletions that distinguish it from DEX0439_001.aa.1 (Ovr107). An alignment of the protein sequences for DEX0439_001.aa.1 (Ovr107) and DEX0439_001.aa.2 is provided in Figure 2.

25 DEX0439_00X.nt.4 (Ovr107v4) encodes the amino acid sequence DEX0439_001.aa.3 which comprises insertions and deletions that distinguish it from DEX0439_001.aa.1 (Ovr107). An alignment of the protein sequences for DEX0439_001.aa.1 (Ovr107) and DEX0439_001.aa.3 is provided in Figure 3.

30 *Ovr107v4 RT-PCR Analysis*

RT-PCR experiments were preformed according to the methods described above in subsection d. Primers were designed unique to Ovr107v4. Results of RT-PCR confirmed the presence of the transcript in a biological sample.

After subcloning, all positively screened clones were sequence verified. DNA
5 sequence verification confirmed Ovr107v4 is distinct from Ovr107.

DEX0439_001.nt.6 Splice Variant

DEX0439_001.nt.6 is a protein encoding sequence supported by 9 ESTs from papillary carcinoma clones and ovarian tumor tissue. This splice variant contains exons that distinguish it from Ovr107. An alignment of the DNA sequences for
10 DEX0439_001.nt.1 (Ovr107) and DEX0439_001.nt.6 is provided in Figure 5.

DEX0439_001.nt.6 encodes the amino acid sequence DEX0439_001.aa.2 which comprises insertions and deletions that distinguish it from DEX0439_001.aa.1 (Ovr107). An alignment of the protein sequences for DEX0439_001.aa.1 (Ovr107) and DEX0439_001.aa.2 is provided in Figure 2.

15 DEX0439_001.nt.6 encodes the amino acid sequence DEX0439_001.aa.3 which comprises insertions and deletions that distinguish it from DEX0439_001.aa.1 (Ovr107). An alignment of the protein sequences for DEX0439_001.aa.1 (Ovr107) and DEX0439_001.aa.3 is provided in Figure 3.

DEX0439_001.nt.8 Splice Variant

20 DEX0439_001.nt.8 is a protein encoding sequence supported by 9 ESTs from papillary carcinoma clones and ovarian tumor tissue. This splice variant contains exons that distinguish it from Ovr107. An alignment of the DNA sequences for DEX0439_001.nt.1 (Ovr107) and DEX0439_001.nt.8 is provided in Figure 6.

DEX0439_001.nt.8 encodes the amino acid sequence DEX0439_001.aa.2 which
25 comprises insertions and deletions that distinguish it from DEX0439_001.aa.1 (Ovr107). An alignment of the protein sequences for DEX0439_001.aa.1 (Ovr107) and DEX0439_001.aa.2 is provided in Figure 2.

DEX0439_001.nt.8 encodes the amino acid sequence DEX0439_001.aa.3 which comprises insertions and deletions that distinguish it from DEX0439_001.aa.1 (Ovr107).
30 An alignment of the protein sequences for DEX0439_001.aa.1 (Ovr107) and DEX0439_001.aa.3 is provided in Figure 3.

DEX0439_001.nt.10 Splice Variant

DEX0439_001.nt.10 is a protein encoding sequence supported by 9 ESTs from papillary carcinoma clones and ovarian tumor tissue. This splice variant contains exons that distinguish it from Ovr107. An alignment of the DNA sequences for DEX0439_001.nt.1 (Ovr107) and DEX0439_001.nt.10 is provided in Figure 7.

5 DEX0439_001.nt.10 encodes the amino acid sequence DEX0439_001.aa.2 which comprises insertions and deletions that distinguish it from DEX0439_001.aa.1 (Ovr107). An alignment of the protein sequences for DEX0439_001.aa.1 (Ovr107) and DEX0439_001.aa.2 is provided in Figure 2.

10 DEX0439_001.nt.10 encodes the amino acid sequence DEX0439_001.aa.3 which comprises insertions and deletions that distinguish it from DEX0439_001.aa.1 (Ovr107). An alignment of the protein sequences for DEX0439_001.aa.1 (Ovr107) and DEX0439_001.aa.3 is provided in Figure 3.

DEX0439_002.nt.1 (Ovr110): Splice Variant DEX0439_002.nt.2 (Ovr110v1)

15 Ovr110 was previously identified wholly or in part as Cancer specific gene (CSG) sequence (clone ID 16656542) in WO 00/12758-A1; Human ovarian carcinoma antigen polynucleotide SEQ ID NO:391 in WO 00/36107-A2; Human B7-H8 gene in WO 02/02587-A1; Human B7-like protein (B7-L) DNA in WO 02/02624-A2 and Breast BS265 gene EST clone 3038129 in US 2002034749-A1 which are herein incorporated by reference.

20 Ovr110 is related to Homo sapiens immune costimulatory protein B7-H4 (B7-H4), mRNA (RefSeq ID: NM_024626.1).

A splice variants has been identified for Ovr110 using the methods described above. It is DEX0439_002.nt.2 (Ovr110v1). This transcript arises from alternative splicing events in the same genomic region as Ovr110 and contains exons encoding amino
25 acid sequences. These amino acid sequences provide proteins to be targeted for the generation of reagents that can be used in the detection and/or treatment of cancer. The nucleotide sequences in these exons can be used as a nucleic acid probe for the diagnosis and/or treatment of cancer.

DEX0439_002.nt.2 (Ovr110v1) Splice Variant

30 Ovr110v1 is a protein encoding sequence supported by 1 EST from prostate tissue. This splice variant contains exons that distinguish it from Ovr110. An alignment of the

DNA sequences for DEX0439_002.nt.1 (Ovr110) and DEX0439_002.nt.2 (Ovr110v1) is provided in Figure 8.

DEX0439_002.nt.2 (Ovr110v1) encodes the amino acid sequence DEX0439_002.aa.2 which comprises insertions and deletions that distinguish it from Ovr110. An alignment of the protein sequences for DEX0439_002.aa.1 (Ovr110) and DEX0439_002.aa.2 is provided in Figure 9.

DEX0439_002.nt.2 (Ovr110v1) encodes the amino acid sequence DEX0439_002.aa.3 which comprises insertions and deletions that distinguish it from Ovr110. An alignment of the protein sequences for DEX0439_002.aa.1 (Ovr110) and DEX0439_002.aa.3 is provided in Figure 10.

Ovr110v1 RT-PCR Analysis

RT-PCR experiments were performed according to the methods described above in subsection d. Primers were designed unique to Ovr110v1. Results of RT-PCR confirmed the presence of the transcript in a biological sample.

After subcloning, all positively screened clones were sequence verified. DNA sequence verification confirmed Ovr110v1 is distinct from Ovr110.

Splice Variant Polypeptide Annotation

The polypeptides of the present invention were analyzed and the following attributes were identified; specifically, epitopes, post translational modifications, signal peptides and transmembrane domains. Antigenicity (Epitope) prediction was performed through the antigenic module in the EMBOSS package. Rice, P., EMBOSS: The European Molecular Biology Open Software Suite, *Trends in Genetics* 16(6): 276-277 (2000). The antigenic module predicts potentially antigenic regions of a protein sequence, using the method of Kolaskar and Tongaonkar. Kolaskar, AS and Tongaonkar, PC., A semi-empirical method for prediction of antigenic determinants on protein antigens, *FEBS Letters* 276: 172-174 (1990). Examples of post-translational modifications (PTMs) and other motifs of the OSPs of this invention are listed below. In addition, antibodies that specifically bind such post-translational modifications may be useful as a diagnostic or as therapeutic. The PTMs and other motifs were predicted by using the ProSite Dictionary of Proteins Sites and Patterns (Bairoch *et al.*, *Nucleic Acids Res.* 25(1):217-221 (1997)), the following motifs, including PTMs, were predicted for the OSPs of the invention. The signal peptides were detected by using the SignalP 2.0, *see Nielsen et al.*, *Protein*

- Engineering* 12, 3-9 (1999). Prediction of transmembrane helices in proteins was performed by the application TMHMM 2.0, "currently the best performing transmembrane prediction program", according to authors (Krogh *et al.*, *Journal of Molecular Biology*, 305(3):567-580, (2001); Moller *et al.*, *Bioinformatics*, 17(7):646-653, (2001);
- 5 Sonnhhammer, *et al.*, *A hidden Markov model for predicting transmembrane helices in protein sequences* in Glasgow, *et al.* Ed. Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology, pages 175-182, Menlo Park, CA, 1998. AAAI Press. The PSORT II program may also be used to predict cellular localizations. Horton *et al.*, *Intelligent Systems for Molecular Biology* 5: 147-152 (1997).
- 10 The table below includes the following sequence annotations: Signal peptide presence; TM (number of membrane domain, topology in orientation and position); Amino acid location and antigenic index (location, AI score, length); PTM and other motifs (type, amino acid residue locations); and functional domains.

DEX AA SEQ ID	Sig P	TMHMM	Antigenicity	PTM	Domain
DEX0439_001.aa.1	N	0 - o	554- 568,1.178; 523- 540,1.153; 327- 334,1.141; 397- 422,1.135; 354- 361,1.134; 58-76,1.133; 130- 149,1.131; 215- 230,1.130; 189- 205,1.127; 120- 128,1.127; 239- 258,1.127; 369- 383,1.122; 573- 579,1.113; 470- 477,1.104; 428- 443,1.091; 154- 160,1.084; 509- 519,1.077; 174-	Amidation 168- 171; Asn_Glycosylation 3-6, 236-239, 371-374; Camp_Phospho_Site 70-73; Ck2_Phospho_Site 17-20, 18-21, 51- 54, 79-82, 111- 114, 190-193, 219-222, 265-268, 274-277, 287-290, 300-303, 468-471, 554-557, 570-573, 579-582; Glycosaminoglycan 38-41; Myristyl 31-36, 84-89, 108-113, 464-469, 511-516, 529-534, 535-540; Pkc_Phospho_Site 38-40, 73-75, 172-174, 183-185, 246-248, 265-267, 376-378, 389-391, 468-470; Tyr_Phospho_Site 187-194;	Proline-rich region; Src homology 3 (SH3) domain; NULL

			182,1.075; 545- 551,1.071; 490- 496,1.055; 298- 304,1.052; 25-31,1.047; 500- 506,1.046; 99- 105,1.045; 40-45,1.041		
DEX0439_001.aa.2	N	0 - o	4-37,1.232; 428- 442,1.178; 397- 414,1.153; 201- 208,1.141; 271- 296,1.135; 228- 235,1.134; 89- 104,1.130; 63-79,1.127; 113- 132,1.127; 243- 257,1.122; 447- 453,1.113; 344- 351,1.104; 302- 317,1.091; 41-56,1.079; 383- 393,1.077; 419- 425,1.071; 364- 370,1.055; 172- 178,1.052; 374- 380,1.046	Asn_Glycosylation 110-113, 245-248; Ck2_Phospho_Site 64-67, 93-96, 139-142, 148-151, 161-164, 174-177, 342-345, 428-431, 444-447, 453-456; Myristyl 338- 343, 385-390, 403-408, 409-414; Pkc_Phospho_Site 57-59, 120-122, 139-141, 250-252, 263-265, 342-344; Tyr_Phospho_Site 61-68;	Proline-rich region; Src homology 3 (SH3) domain; NULL
DEX0439_001.aa.3	N	0 - o	4-41,1.232; 205- 212,1.141; 275- 300,1.135; 232- 239,1.134; 93- 108,1.130; 67-83,1.127; 117- 136,1.127;	Amidation 406- 409; Asn_Glycosylation 114-117, 249-252; Ck2_Phospho_Site 68-71, 97-100, 143-146, 152-155, 165-168, 178-181, 346-349; Myristyl 342-347, 389-394; Pkc_Phospho_Site 61-63, 124-126,	Proline-rich region; Src homology 3 (SH3) domain

			247- 261,1.122; 348- 355,1.104; 306- 321,1.091; 45-60,1.079; 387- 397,1.077; 368- 374,1.055; 176- 182,1.052; 378- 384,1.046	143-145, 254-256, 267-269, 346-348, 406-408; Tyr_Phospho_Site 65-72;	
DEX0439_002.aa.1	Y	1 - 17- 29o	207- 217,1.215; 248- 279,1.179; 176- 186,1.162; 73-83,1.160; 4-30,1.146; 100- 123,1.136; 127- 134,1.125; 61-71,1.119; 36-45,1.103; 152- 160,1.069; 53-59,1.066; 165- 171,1.060	Asn_Glycosylation 117-120, 165-168, 195-198, 201-204, 210-213, 221-224, 225-228; Camp_Phospho_Site 251-254; Ck2_Phospho_Site 96-99, 156-159, 188-191, 202-205, 246-249; Myristyl 57-62, 131-136, 193-198; Pkc_Phospho_Site 37-39, 119-121, 132-134, 139-141, 170-172, 212-214;	Immunoglobulin and major histocompatibility complex domain
DEX0439_002.aa.2	y	1 - 17- 29o	41-56,1.170; 4-39,1.146		
DEX0439_002.aa.3	N	0 - o	177- 187,1.215; 218- 249,1.179; 146- 156,1.162; 43-53,1.160; 70-93,1.136; 97- 104,1.125; 31-41,1.119; 6-15,1.103; 23-29,1.100; 122- 130,1.069; 135- 141,1.060	Asn_Glycosylation 87-90, 135-138, 165-168, 171-174, 180-183, 191-194, 195-198; Camp_Phospho_Site 221-224; Ck2_Phospho_Site 66-69, 126-129, 158-161, 172-175, 216-219; Myristyl 27-32, 101-106, 163-168; Pkc_Phospho_Site 89-91, 102-104, 109-111, 140-142, 182-184;	Immunoglobulin and major histocompatibility complex domain; Immunoglobulin subtype

Altogether, splice variant sequence analysis, EST support, SAGE tag data, RT-PCR data, real time Quantitative PCR data, Secretion assay data, and protein annotation

are indicative of SEQ ID NO: 1-8 and encoded protein SEQ ID NO: 9-14 being a diagnostic marker and/or a therapeutic target for cancer.

Example 2: Gene Expression Analysis

Custom Microarray Experiment—Ovarian Cancer

5 Custom oligonucleotide microarrays were provided by Agilent Technologies, Inc. (Palo Alto, CA). The microarrays were fabricated by Agilent using their technology for the *in-situ* synthesis of 60mer oligonucleotides. Hughes *et al.*, *Nature Biotechnology* 19:342-347(2001). The 60mer microarray probes were designed by Agilent, from gene sequences provided by diaDexus, using Agilent proprietary algorithms. Whenever
10 possible two different 60mers were designed for each gene of interest.

All microarray experiments were two-color experiments and were preformed using Agilent-recommended protocols and reagents. Briefly, each microarray was hybridized with cRNAs synthesized from polyA+ RNA, isolated from cancer and normal tissues, labeled with fluorescent dyes Cyanine3 (Cy3) and Cyanine5 (Cy5) (NEN
15 Life Science Products, Inc., Boston, MA) using a linear amplification method (Agilent). In each experiment the experimental sample was polyA+ RNA isolated from cancer tissue from a single individual and the reference sample was a pool of polyA+ RNA isolated from normal tissues of the same organ as the cancerous tissue (*i.e.* normal ovarian tissue in experiments with ovarian cancer samples). Hybridizations were carried out at 60°C,
20 overnight using Agilent *in-situ* hybridization buffer. Following washing, arrays were scanned with a GenePix 4000B Microarray Scanner (Axon Instruments, Inc., Union City, CA). The resulting images were analyzed with GenePix Pro 3.0 Microarray Acquisition and Analysis Software (Axon).

Two different chip designs were evaluated with overlapping sets of a total of 19
25 samples, comparing the expression patterns of ovarian cancer derived total RNA to total RNA isolated from a pool of 9 normal ovarian tissues. For Chip 2, all 19 samples (14 invasive carcinomas, 5 low malignant potential samples were analyzed and for Chip 1, a subset of 17 of these samples (13 invasive carcinomas, 4 low malignant potential samples) were assessed.

30 Data normalization and expression profiling were done with Expressionist software from GeneData Inc. (Daly City, CA/Basel, Switzerland). Gene expression analysis was performed using only experiments that meet certain quality criteria. The

quality criteria that experiments must meet are a combination of evaluations performed by the Expressionist software and evaluations performed manually using raw and normalized data. To evaluate raw data quality, detection limits (the mean signal for a replicated negative control + 2 Standard Deviations (SD)) for each channel were calculated. The
5 detection limit is a measure of non-specific hybridization. Acceptable detection limits were defined for each dye (80 for Cy5 and 150 for Cy3). Arrays with poor detection limits in one or both channels were not analyzed and the experiments were repeated. To evaluate normalized data quality, positive control elements included in the array were utilized. These array features should have a mean ratio of 1 (no differential expression).
10 If these features have a mean ratio of greater than 1.5-fold up or down, the experiments were not analyzed further and were repeated. In addition to traditional scatter plots demonstrating the distribution of signal in each experiment, the Expressionist software also has minimum thresholding criteria that employ user defined parameters to identify quality data. These thresholds include two distinct quality measurements: 1) minimum
15 area percentage, which is a measure of the integrity of each spot and 2) signal to noise ratio, which ensures that the signal being measured is significantly above any background (nonspecific) signal present. Only those features that meet the threshold criteria were included in the filtering and analyses carried out by Expressionist. The thresholding settings employed require a minimum area percentage of 60% [(% pixels > background +
20 2SD)-(% pixels saturated)], and a minimum signal to noise ratio of 2.0 in both channels. By these criteria, very low expressors, saturated features and spots with abnormally high local background were not included in analysis.

Relative expression data was collected from Expressionist based on filtering and clustering analyses. Up-regulated genes were identified using criteria for the percentage
25 of experiments in which the gene is up-regulated. In general, up-regulation in ~30% of samples tested was used as a cutoff for filtering.

Microarray results include information about the sequence itself (DEX ID, OligoID), results obtained for all ("ALL") ovarian cancer samples, invasive carcinomas ("INV") and low malignant potential ("LMP") samples. '%up' indicates the percentage of
30 all experiments in which up-regulation of at least 2-fold was observed (n=17 for Chip 1, n=19 for Chip 2), '%valid up' indicates the percentage of experiments with valid expression values in which up-regulation of at least 2-fold was observed.

Example 3: Protein Expression

The OSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the OSNA is subcloned in pET-21d for expression in *E. coli*. In addition to the OSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH₂-terminus of the coding sequence of OSNA, and six histidines, flanking the COOH-terminus of the coding sequence of OSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of OSP is achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that are separated from total cell lysate were incubated with a nickel chelating resin. The column is packed and washed with five column volumes of wash buffer. OSP is eluted stepwise with various concentration imidazole buffers.

Example 4: Fusion Proteins

The human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e.g.*, WO 96/34891.

Example 5: Production of an Antibody from a Polypeptide

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. See, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1-8. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C;

and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al.*, *Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al.*, *Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The
5 intron-exon borders of selected exons are also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al.*, *Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected
10 individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Mannheim), and FISH is performed as described in Johnson *et al.*, *Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1
15 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and
20 variable excitation wavelength filters. Johnson (1991). Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

25 **Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample**

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal
30 and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial

dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 :l of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three
5 times with deionized or distilled water to remove unbound conjugate. 75 :l of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are
10 plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion
15 consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

20 As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1, $\mu\text{g/kg/day}$ to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given
25 continuously, the secreted polypeptide is typically administered at a dose rate of about 1 $\mu\text{g/kg/hour}$ to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the
30 desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally,

topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which
5 include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules. Sustained-
10 release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481, the contents of which are hereby incorporated by reference herein in their entirety), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate
15 (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese
20 Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324, the contents of which are hereby incorporated by reference herein in their entirety. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

25 For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

30 For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is

shaped into the desired formulation. Preferably, the carrier is a parenteral carrier, more preferably, a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as

5 liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic

10 acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such

15 as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers

20 will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a

25 stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture

30 is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions

of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may
5 be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form.
10 Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose
15 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense or RNAi technology are used to inhibit production of a polypeptide of
20 the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the
25 treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a
30 subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a

tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 3. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

- 5 The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 12: Method of Treatment Using Gene Therapy-In Vivo

- Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the
10 introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

- The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in
15 the art, see, for example, Tabata H. *et al. Cardiovasc. Res.* 35 (3): 470-479 (1997); Chao J *et al. Pharmacol. Res.* 35 (6): 517-522 (1997); Wolff J. A. *Neuromuscul. Disord.* 7 (5): 314-318 (1997), Schwartz B. *et al. Gene Ther.* 3 (5): 405-411 (1996); and Tsurumi Y. *et al. Circulation* 94 (12): 3281-3290 (1996); W0 90/11092, W0 98/11779; U. S. Patent No. 5,693,622; 5,705,151; 5,580,859, the contents of which are hereby incorporated by
20 reference herein in their entirety.

- The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, ovarian, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

- 25 The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. *et al. Ann. NY Acad. Sci.* 772: 126-139 (1995) and Abdallah B. *et al. Biol. Cell* 85 (1): 1-7 (1995)) which
30 can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, ovarian, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. In vivo muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 $\mu\text{g/kg}$ body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg . Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also

be used, such as, inhalation of an aerosol formulation particularly for delivery to ovarians or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

5 The dose response effects of injected polynucleotide in muscle in vivo is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps
10 muscles of mice are then injected with various amounts of the template DNA.

 Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute,
15 approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

 After an appropriate incubation time (e.g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual
20 quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

25 The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

Example 13: Transgenic Animals

 The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea
30 pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific

embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (I. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver et al., *Biotechnology* 11: 1263-1270 (1993); Wright et al., *Biotechnology* 9: 830-834 (1991); and U. S. Pat. No. 4,873,191, the contents of which is hereby incorporated by reference herein in its entirety); retrovirus mediated gene transfer into germ lines (Van der Putten et al., *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, *Mol Cell. Biol.* 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., *Science* 259: 1745 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., *Cell* 57: 717-723 (1989). For a review of such techniques, see Gordon, "Transgenic Animals," *Intl. Rev. Cytol.* 115: 171-229 (1989).

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., *Nature* 380: 64-66 (1996); Wilmut et al., *Nature* 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, I. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e.g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of

integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., *Science* 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., *Nature* 317: 230-234 (1985); Thomas & Capecchi, *Cell* 51: 503512 (1987); Thompson et al., *Cell* 5: 313-321 (1989)) Alternatively, RNAi technology may be used. For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention in vivo. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, *supra*). However, this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site in vivo using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient in vivo. Such cells may be obtained from the patient (i.e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered in vitro using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e.g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be
5 introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent No. 5,399,349; and
10 Mulligan & Wilson, U. S. Patent No. 5,460,959, the contents of which are hereby incorporated by reference herein in their entirety).

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the
15 cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with
20 aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other
25 than the described embodiments, which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

SEQUENCES

>DEX0439_001.nt.1

AGAGCAAGGAAGGGCAGGGGACCTGGGAAGGAAGTTCTGGAAGGCAGTGGGGTTTGAGAT
TGGACCCAGGGTCAAGATAGAACATGAAGGTGGGATGAGGACATGAACAGAACATGGCCA
5 AGAAGGATCTGGGGGAGCAGCCAGGACGAGGCGGAGCTGATCCGAGAGGACATCCAGGGG
GCTCTGCACAATTACCGCTCGGGCCGCGGGGAGCGCAGGGCGGGCGCTCAGGGCCACG
CAGGAGGAGTTGCAGCGCGACCGCTCGCCCCGCCGTGAGACCCCGCCCCCTGCAGCGCCGC
CCGTCAGTCCGCGCAGTGATCAGCACCGTAGAGCGGGCGCGGGCCGCGGACGACCCAG
GCGAAGCCCATTCCCGAGGCAGAGGAGGCGCAGAGGCTGAGCCGGTGGGGACCTCGAGC
10 AACGCTGACTCGGCCTCCCCGACCTGGGTCCCCGGGGTCTGACCTGGCGGTTCTGCAG
GCGGAGCGGGAAGTGGACATCCTGAACCACGTGTTGACGACGTAGAGAGCTTTGTATCG
AGGTGTCAGAAGTCGGCGGAGGCGGCCAGGGTGTGGAGCACCGGGAACGCGGCCGAGG
AGCCGGCGCCGGGCGGCTGGGGAGGGCTTGCTGACGCTGCGGGCCAAGCCGCCCTCGGAG
GCCGAGTACACCGAGTGTGTCAGAAGTACGCTTCAGCCTGCTGGCCCGGCTG
15 CGCGGAACATCGCCGACCCCTCCTCTCCGGAGCTGTTGCACTTCCTTTTCGGGCTCTG
CAGATGATTGTGAACACGTGCGGGGGGCGGGAGTTCCGAGCAGTGTGCGGCGGCCGAT
CTGACATCGGATGCCGTGGCGCTGCTGCGGGACAACGCTCACTCCACGTGAAAACGAGCTC
TGGACCTCGCTGGGGGACTCGTGGACCCGCCCCGGGCTGGAGCTGTCCCCGGAGGAGGGA
CCCCATACAGACCCGAGTTCTTACGGCTGGGAGCCGCCGCTCACTGACCCGAGAGC
20 CGCGCTGGGAGGACCCAGTTGAGAAACAGCTACAGCACGAGCGGAGGCGCCGCGCAAA
AGCGCCCCCAGGTGCTGTCAATGGTCAACGAGACTTGGAGCCAGAATCTGAGCCTCAG
CTGGAGTCAGAGACAGCAGGAAATGGGTCTGTGTAATTATGACTTCAGGCCCGCAAC
AGCAGTGAGCTGTGGTCAAGCAGCGGGACGTACTGGAGGTCTGGATGACAGTCGTAAG
TGGTGGAAAGTTTCGGGACCCAGCGGGGCGAGGAGGATATGTGCCCTACAACATCCGACA
25 CCTACCCCGGACCCCGGCTGCACCACAGCCAAAGCCCTGCCCGAGCCTGAACAGCACT
CCTCCTCCACCACCAGCCCCAGCCCCGGCCCCACCTCCAGCTCTGGCTCGGCCCGCTGG
GACAGGCCCCGCTGGGACAGCTGCGATAGCCTCAACGGCTTGGACCCAGCGAGAGGAG
AAATTCTCCAGATGCTCATCGTCAACGAGGAAGTGCAGGCGCGCCTGGCCAGGGCCGC
TCGGGACCGAGCCGCGCAGTCCAGGGCCCCCGCGCCCCGGAACCGCAGCTCAGCCGGGC
30 TCGGACGCTCCGAGGTCCGCGCCTGGCTGCAGGCCAAGGGCTTTAGCTCCGGGACCGTG
GACGCGCTGGGTGTGCTGACCGGGGCGCAGCTTTTCTCGCTGCAGAGGGAGGAGTCCGG
GCGGTGAGCCCCGAGGAGGGGGCACGTGTGTACAGCCAGGTACCCGTGCAGCGCTCGCTG
CTGGAGGACAAAGAGAAAGTGTGACAGCTGGAGGCACTGATGGAGAAGCAAAAGAAAG
GTGGAGGCGAGGTGGAATGGAGGTCAATTGACCTGCCAGGCGCCCTTCGCAAGAGTG
35 ACGAGGCCCGTGGGAGAACGGACTCCTCAGACTCTCCCCAATAGCGGAAGTCGATCTTC
TGAAGGATGGCCAATCTGCTCCGGCCCTGGTCTTCCCCCATCCCGGTGGACAGACTTAAC
GATCCTTGCTGCAGTCCCTCCGGAGAGGATCTGGACTGGCTGGGAGTGGGGAGGGCGTGG
AGCAGTCTACGGAAAGCGCTAGCAGACCCCGAGAGGGTGCAGTGGAGCCCTGAGCAATT
GTAATATGCGGCCAGCCTATAAACAGCCTCCGTGCTTAGCAAAAAAAAAAAAAAAAAA
40 A

>DEX0439_001.nt.2

GATCTCTTCAAATGTCCCCGCTCTCCCCAGGCTCTCCCCCTCCCGCCACTTGCCAGGGCT
GACCTCACCGCCATCTTAACCGGGTGCCACCTCTCTGCTGCTGCTGGTGTGGCCCCG
45 CGTCCCCATCGCCGCGCCCGTCTGCTCCCCCTCAGAGGGCTTGCTGACGCTGCGGGCCAAG
CCGCCCTCGGAGGCCGAGTACACCGACGTGCTGCAGAAGATCAAGTACGCCTTCAGCCTG
CTGGCCCGGCTGCGCGGAACATCGCCGACCCCTCCTCTCCGGAGCTGTTGCACTTCCTT
TTCGGGCTCTGCAGATGATTGTGAACACGTGCGGGGGGCGGAGTTTCGCGAGCAGTGTG
CGCGGCGCGCATCTGACATCGGATGCCGTGGCGCTGCTGCGGGACAACGTCACTCCACGT
50 GAAAACGAGCTCTGGACCTCGCTGGGGGACTCGTGGACCCGCCCCGGGCTGGAGCTGTCC
CCGGAGGAGGGACCCCATACAGACCCGAGTTCTTCAGCGGCTGGGAGCCGCCGCTCACT
GACCCGCGAGCCGCGCCTGGGAGGACCCAGTTGAGAAACAGCTACAGCACGAGCGGAGG
CGCCGGCAGCAAGCGCCCCCAGGTGCTGTCAATGGTCAACGAGACTTGGAGCCAGAA
TCTGAGCCTCAGCTGGAGTCAGAGACAGCAGGAAATGGGTCTGTGTAATTATGACTTC
CAGGCCCGCAACAGCAGTGAAGTGTGCGGTCAAGCAGCGGGACGTACTGGAGGTCTGGAT
55 GACAGTCGTAAGTGGTGGAAAGTTTCGGGACCCAGCGGGGAGGAGGATATGTGCCCTAC
AACATCCTGACACCCCTACCCCGGACCCCGGCTGCACCACAGCCAAAGCCCTGCCCGCAGC
CTGAACGCACTCCTCCTCCACCAGCCCCAGCCCCGGCCCCACCTCCAGCTCTGGCT
CGGCCCGCTGGGACAGGCCCGCTGGGACAGCTGCGATAGCCTCAACGGCTTGGACCCC
AGCGAGAAGGAGAAATTCTCCAGATGCTCATCGTCAACGAGGAAGTGCAGGCGCGCCTG

GCCCAGGGCCGCTCGGGACCGAGCCGCGCAGTCCCAGGGCCCCGCGCCCCGGAACCGCAG
CTCAGCCCGGGCTCGGACGCCTCCGAGGTCCGCGCCTGGCTGCAGGCCAAGGGCTTTAGC
TCCGGAAGGAGGAGCTCGGGCGGTGAGCCCCGAGGAGGGGGCACGTGTGTACAGCCAGG
5 TCACCGTGCAGCGCTCGCTGCTGGAGGACAAAGAGAAAGTGTACAGAGCTGGAGGCAGTGA
TGGAGAAGCAAAAGAAGAAGGTGGAAGGCGAGGTGGAAATGGAGGTCATTTGACCTGCCA
GGCGCCCTTCGCAAGAGTGACGAGGCCCGTGGGAGAACGGACTCCTCAGACTCTCCCC
AATAGCGGAAGTCGATCTTCTGAAGGATGGCCAATCTGCTCCGGCCCTGGTCTTCCCCCA
TCCCGGTGGACAGACTTAACGATCCTTGCTGCAGTCCCTCCGGAGAGGATCTGGACTGGC
10 TGGAGTGGGGAGGCGTGGAGACAGTCTACGGAAGCGCTAGCAGACCCCCGAGAGGGT
GCAGTGGAGCCCTGAGCATTGTAATATGCGGCCAGCCTATAAACAGCCTCCGTGCTTAG
CAG
>DEX0439_001.nt.4
GATCTCTTCCAAATGTCCCCGCTCTCCCCAGGCTCTCCCCTCCCGCCACTTGCCAGGGCT
GACCTCACCGCCATCTTAACCGGGTGTCCACCTCTCTCTGCCTGCCGTGGTGTGGCCCCG
15 CGTCCCCATCGCCGCGCCCGTCTGCTCCCCTCAGAGGGCTTGCTGACGCTGCGGGCCAAG
CCGCCCTCGGAGGCCGAGTACACCGACGTGCTGCAGAAGATCAAGTACGCCTTCAGCCTG
CTGGCCCCGCTGCGCGCAACATCGCCGACCCCTCCTCTCCGGAGCTGTTGCACTTCCTT
TTCGGGCCTCTGCAGATGATTGTGAACACGTGCGGGGGGCGGGAGTTCGCGAGCAGTGTG
CGCGCGCCGATCTGACATCGGATGCCGTGGCGCTGCTGCGGGACAACGTCACTCCACGT
20 GAAAACGAGCTCTGGACCTCGCTGGGGGACTCGTGGACCCGCCCCGGGCTGGAGCTGTCC
CCGGAGGAGGCCCCCATACAGACCCGAGTTCTTCAGCGGTGGGAGCCGCGCTCACT
GACCCGCAGAGCCGCGCTGGGAGGACCCAGTTGAGAAACAGCTACAGCACGAGCGGAGG
CGCCGGCAGCAAAGCGCCCCCAGGTCGCTGTCAATGGTCACCGAGACTTGGAGCCAGAA
TCTGAGCCTCAGCTGGAGTCAGAGACAGCAGGAAAATGGGTCTGTGTAATTATGACTTC
25 CAGGCCCGCAACAGCAGTGAGCTGTGCTCAAGCAGCGGGACGTACTGGAGGTCTGGAT
GACAGTCGTAAGTGGTGGAAAGTTTCGGGACCCAGCGGGGCAGGAGGGATATGTGCCCTAC
AACATCCTGACACCCCTACCCCGGACCCCGGCTGCACCACAGCCAAAGCCCTGCCCGCAGC
CTGAACAGCACTCCTCCTCCACCACCAGCCCGCCCGGCCCCACCTCCAGCTCTGGCT
CGGCCCCGCTGGGACAGGCCCGCTGGGACAGCTGCGATAGCCTCAACGGCTTGGACCCC
30 AGCGAGAAGGAGAAATTCTCCAGATGCTCATCGTCAACGAGGAAGTGCAGGCGCGCCTG
GCCAGGGCCGCTCGGACCGAGCCGCGCAGTCCAGGGCCCCCGCGCCCCGGAACCGCAG
CTCAGCCCGGGCTCGGACGCCTCCGAGGTCCGCGCCTGGCTGCAGGCCAAGGGCTTTAGC
TCCGGGACCGTGGACGCGCTGGGTGTGCTGACCGGGGCGCAGCTTTCTCGCTGCAGAAG
GAGGAGCTCGGGCGGTGAGCCCCGAGGAGGGGGCACGTGTGTACAGCCAGGTACCCGTG
35 CAGCGCTCGCTGCTGGAGGACAAAGAGAAAGTGTACAGAGCTGGAGGCAGTGATGGAGAAG
CAAAAGAAGAAGGTGGAAGGCGAGGTGGAATGGAGGTCATTTGACCTGCCAGGCGCCCT
TCGCAAGAGTGACGAGGCCCGTGGGAGAACGGACTCCTCAGACTCTCCCCAATAGCGG
AAGTCGATCTTCTGAAGGATGGCCAATCTGCTCCGGCCCTGGTCTTCCCCCATCCCGGTG
GACAGCTTAACGATCCTTGCTGCAGTCCCTCCGGAGAGGATCTGGACTGGCTGGGAGTG
40 GGGAGGGCGTGGAGACAGTCTACGGAAGCGCTAGCAGACCCCGAGAGGGTGCAGTGA
GCCCTGAGCATTGTAATATGCGGCCAGCCTATAAACAGCCTCCGTGCTTAGCAG
>DEX0439_001.nt.6
GATCTCTTCCAAATGTCCCCGCTCTCCCCAGGCTCTCCCCTCCCGCCACTTGCCAGGGCT
GACCTCACCGCCATCTTAACCGGGTGTCCACCTCTCTCTGCCTGCCGTGGTGTGGCCCCG
45 CGTCCCCATCGCCGCGCCCGTCTGCTCCCCTCAGAGGGCTTGCTGACGCTGCGGGCCAAG
CCGCCCTCGGAGGCCGAGTACACCGACGTGCTGCAGAAGATCAAGTACGCCTTCAGCCTG
CTGGCCCCGCTGCGCGCAACATCGCCGACCCCTCCTCTCCGGAGCTGTTGCACTTCCTT
TTCGGGCCTCTGCAGATGATTGTGAACACGTGCGGGGGGCGGGAGTTTCGCGAGCAGTGTG
50 CGGCGGCCGATCTGACATCGGATGCCGTGGCGCTGCTGCGGGACAACGTCACTCCACGT
GAAAACGAGCTCTGGACCTCGCTGGGGGACTCGTGGACCCGCCCCGGGCTGGAGCTGTCC
CCGGAGGAGGAGCCCCCATACAGACCCGAGTTCTTCAGCGGTGGGAGCCGCGGTCAGT
GACCCGCAGAGCCGCGCCTGGGAGGACCCAGTTGAGAAACAGCTACAGCACGAGCGGAGG
CGCCGGCAGCAAAGCGCCCCCAGGTCGCTGTCAATGGTCACCGAGACTTGGAGCCAGAA
TCTGAGCCTCAGCTGGAGTCAGAGACAGCAGGAAAATGGGTCTGTGTAATTATGACTTC
55 CAGGCCCGCAACAGCAGTGAGCTGTGCTCAAGCAGCGGGACGTACTGGAGGTCTGGAT
GACAGTCGTAAGTGGTGGAAAGTTTCGGGACCCAGCGGGGCAGGAGGGATATGTGCCCTAC
AACATCCTGACACCCCTACCCCGGACCCCGGCTGCACCACAGCCAAAGCCCTGCCCGCAGC
CTGAACAGCACTCCTCCTCCACCACCAGCCCCAGCCCCGGCCCCACCTCCAGCTCTGGCT
CGGCCCCGCTGGGACAGGCCCGCTGGGACAGCTGCGATAGCCTCAACGGCTTGGACCCC
60 AGCGAGAAGGAGAAATTCTCCAGATGCTCATCGTCAACGAGGAAGTGCAGGCGCGCCTG

GCCCAGGGCCGCTCGGGACCGAGCCGCGCAGTCCCAGGGCCCCGCGCCCCGGAACCGCAG
CTCAGCCCCGGGCTCGGACGCCTCCGAGGTCCGCGCCTGGCTGCAGGCCAAGGGCTTTAGC
TCCGGGACCGTGGACGCGCTGGGTGTGCTGACCGGGCGCAGCTTTTCTCGCTGCAGAAG
GAGGAGCTGCGGGCGGTGAGCCCCGAGGAGGGGGCACGTGTGTACAGCCAGGTACCCGTG
5 CAGCGCTCGCTGCTGGAGGACAAAGAGAAAGTGTGAGAGCTGGAGGCAGTGATGGAGAAG
CAAAAGAAGAAGGTGGAAGGCGAGGTGGAAATGGAGGTCAATTTGACCTGCCAGGCGCCCT
TCGCAAAGAGTGACGAGGCCCCGTGGGAGAACGGACTCCTCAGACTCTCCCCAATAGCGG
AAGTCGATCTTCTGAAGGATGGCCAATCTGCTCCGGCCCTGGTCTTCCCCCATCCCGGTG
GACAGACTTAACGATCCTTGTGTCAGTCCCTCCGGAGAGGATCTGGACTGGCTGGGAGTG
10 GGGAGGGCGTGGAGACAGTCTACGAAAGCGCTAGCAGACCCCCGAGAGGGTGCAGTGGA
GCCCTGAGCATTGTAATATGCGGCCCCAGCCTATAAACAGCCTCCGTGCTTAGCAGAAAAA
AAACATACAAAAAT
AAAAGATAGTCAACAAACAAAATAAGAACTATAGATAATATAAAAATGAAAATAAAAA
AGAGATGGGTGGGGCCCTTGTCTTTACTCTCTCCCTCTGGAGTGGGCACACTATATTAT
15 TTCGCCCTCCCCCTCTTTTTTTGTATGAGAGGGCTCTTTTA
>DEX0439_001.nt.8
GATCTCTTCCAAATGTCCCCGCTCTCCCCAGGCTCTCCCTCCCCGCACTTGCCAGGGCT
GACCTCACCGCCATCTTAACCGGGTGTCCACCTCTCTCTGCCTGCCTGGTGTGCGCCCG
CGTCCCCATCGCCGCGCCCGTCTGCTCCCTCAGAGGGCTTGCTGACGCTGCGGGCCAAG
20 CCGCCCTCGGAGGCCGAGTACACCGAGCTGCTGCAGAAGATCAAGTACGCCTTCAGCCTG
CTGGCCCGGTGCGCGGCAACATCGCCGACCCCTCCTCTCCGGAGCTGTTGCACTTCCTT
TTCGGGCCCTCTGCAGATGATTGTGAACACGTGCGGGGGGCGGAGTTCGCGAGCAGTGTG
CGGCGGCCGATCTGACATCGGATGCCGTGGCGCTGCTGCGGACAACGTCACTCCACGT
GAAAACGAGCTCTGGACCTCGCTGGGGGACTCGTGGACCCGCCCGGGCTGGAGCTGTCC
25 CCGGAGGAGGGACCCCCATACAGACCCGAGTTCTTCAGCGGCTGGGAGCCGCGCGTCACT
GACCCGAGAGCCGCGCCTGGGAGGACCCAGTTGAGAAACAGCTACAGCACGAGCGGAGG
CGCCGGCAGCAAGCGCCCCCAGGTGCTGCAATGGTCACCGAGACTTGAGCCAGAA
TCTGAGCCTCAGCTGGAGTCAGAGACAGCAGGAAAATGGGTCTGTGTAATTATGACTTC
CAGGCCCCGCAACAGCAGTGAGCTGTGCGTCAAGCAGCGGGACGTACTGGAGGTCTGGAT
30 GACAGTCGTAAGTGGTGGAAAGGTTCCGGACCCAGCGGGGAGGAGGGATATGTGCCCTAC
AACATCCTGACACCTACCCCGGACCCCGGCTGCACCACAGCCAAAGCCCTGCCCGCAGC
CTGAACGACACTCCTCCTCCACCACAGCCCCAGCCCCGGCCCCACCTCCAGCTCTGGGT
CGGCCCCGCTGGGACAGGCCCCGCTGGGACAGCTGCGATAGCCTCAACGGCTTGAGCCCC
AGCGAGAAGGAGAAATCTCCAGATGCTCATCGTCAACGAGGAACTGCAGGCGCGCCTG
35 GCCCAGGGCCGCTCGGGACCGAGCCGCGCAGTCCCAGGGCCCCGCGCCCCGGAACCGCAG
CTCAGCCCCGGCTCGGACGCTCCTCCAGGTCCGCGCCTGGCTGCAGGCCAAGGGCTTTAGC
TCCGGGACCGTGGACGCGCTGGGTGTGCTGACCGGGGCGAGCTTTCTCGCTGCAGAAG
GAGGAGCTGCGGGCGGTGAGCCCCGAGGAGGGGGCACGTGTGTACAGCCAGGTACCCGTG
CAGCGCTCGCTGCTGGAGGACAAAGAGAAAGTGTGAGAGCTGGAGGCAGTGATGGAGAAG
40 CAAAAGAAGAAGGTGGAAGGCGAGGTGGAATGGAGGTCAATTTGACCTGCCAGGCGCCCT
TCGCAAAGAGTGACGAGGCCCGTGGGAGAACGGACTCCTCAGACTCTCCCCAATAGCGG
AAGTCGATCTTCTGAAGGATGGCCAATCTGCTCCGGCCCTGGTCTTCCCCCATCCCGGTG
GACAGACTTAACGATCCTTGTGTCAGTCCCTCCGGAGAGGATCTGGACTGGCTGGGAGTG
GGGAGGGCGTGGAGACAGTCTACGGAAGCGCTAGCAGACCCCCGAGAGGGTGCAGTGGA
45 GCCCTGAGCATTGTAATATGCGGCCCCAGCCTATAAACAGCCTCCGTGCTTAGCAGAAAAA
SAAACACATCAACCAACAAACGTTTGGGGTATTCCATGGCCAATACCGTTGTTCCCGT
GTGTGAACATTGTTATTTTCAGCTCACATTTCCACAGTATTGGAACAACACATCATACCA
CACACACACAGAACCAATCGAGATATATAAACCAATGCACACTCAAACACCTAAT
>DEX0439_001.nt.10
50 GATCTCTTCCAAATGTCCCCGCTCTCCCCAGGCTCTCCCTCCCCGCACTTGCCAGGGCT
GACCTCACCGCCATCTTAACCGGGTGTCCACCTCTCTCTGCCTGCCTGGTGTGCGCCCG
CGTCCCCATCGCCGCGCCCGTCTGCTCCCTCAGAGGGCTTGCTGACGCTGCGGGCCAAG
CCGCCCTCGGAGGCCGAGTACACCGACGTGCTGCAGAAGATCAAGTACGCCTTCAGCCTG
CTGGCCCGGTGCGCGGCAACATCGCCGACCCCTCCTCTCCGGAGCTGTTGCACTTCCTT
55 TTCGGGCCCTCTGCAGATGATTGTGAACACGTGCGGGGGGCGGAGTTCGCGAGCAGTGTG
CGGCGCCGATCTGACATCGGATGCCGTGGCGCTGCTGCGGGACAACGTCACTCCACGT
GAAAACGAGCTCTGGACCTCGCTGGGGGACTCGTGGACCCGCCCGGGCTGGAGCTGTCC
CCGGAGGAGGGACCCCCATACAGACCCGAGTTCTTCAGCGGCTGGGAGCCGCGCGTCACT
GACCCGAGAGCCGCGCCTGGGAGGACCCAGTTGAGAAACAGCTACAGCACGAGCGGAGG
60 CGCCGGCAGCAAGCGCCCCCAGGTGCTGTCAATGGTCACCGAGACTTGAGGCCAGAA

TCTGAGCCTCAGCTGGAGTCAGAGACAGCAGGAAAATGGGTCTGTGTAATTATGACTTC
CAGGCCCCGAACAGCAGTGTGCTGGTCAAGCAGCGGGACGTACTGGAGGTCTGGAT
GACAGTCGTAAGTGGTGAAGGTTCTGGGACCCAGCGGGGACAGGAGGATATGTGCCCTAC
AACATCTTGACACCCCTACCCCGGACCCCGGCTGCACCAAGCCAAAGCCCTGCCCGCAGC
5 CTGAACAGCACTCCTCTCCACCAGCCCCAGCCCCGGCCCCACCTCCAGCTCTGGCT
CGGCCCCGCTGGGACAGCCCCGCTGGGACAGCTGCGATAGCCTCAACGGCTTGGACCCC
AGCGAGAAGGAGAAATTCTCCAGATGCTCATCGTCAACGAGGAAGTGCAGGCGCGCCTG
GCCAGGCGCGCTCGGGACCGAGCCGCGCAGTCCCAGGGCCCCGCGCCCCGGAACCGCAG
10 CTCAGCCCCGGGCTCGGACGCCTCCGAGGTCCGCGCTGGCTGCAGGCCAAGGGCTTTAGC
TCCGGGACCGTGGACGCGCTGGGTGTGCTGACCGGGGCGCAGCTTTTCTCGCTGCAGAAG
GAGGAGCTGCGGGCGGTGAGCCCCGAGGAGGGGCGACGTGTGTACAGCCAGGTACCCGTG
CAGCGCTCGCTGCTGGAGGACAAAGAGAAAGTGTGAGAGCTGGAGGCACTGATGGAGAAG
CAAAAGAAGAAGGTGGAGGCGAGGTGGAAATGGAGGTCATTTGACCTGCCAGGCGCCCT
15 TCGCAAAGAGTGACGAGGCCCGTGGGAGAACGGACTCCTCAGACTCTCCCCAATAGCGG
AAGTCGATCTTCTGAAGGATGGCCAATCTGCTCCGGCCCTGGTCTTCCCCCATCCCGGTG
GACAGACTTAACGATCCTTGCTGCAGTCCCTCCGAGAGGATCTGGACTGGCTGGGAGTG
GGGAGGGCGTGGAGACAGTCTACGGAAAGCGCTAGCAGACCCCCGAGAGGGTGCAGTGGA
GCCCTGAGCATTGTAATATGCGGCCAGCCTATAAACAGCCTCCGTGCTTAGCAG
>DEX0439_002.nt.1
20 GAGTACCAAGGAAGGCAGCGGCAGCTCCACTCAGCCAGTACCCAGATACGCTGGGAACC
TTCCCCAGCCATGGCTTCCCTGGGGCAGATCCTCTTCTGGAGCATAATTAGCATCATCAT
TATCTGGCTGGAGCAATTGCACTCATCATTGGCTTTGGTATTTAGGGAGACACTCCAT
CACAGTCACTACTGTGCGCTCAGCTGGGAACATTGGGGAGGATGGAATCCAGAGCTGCAC
25 TTTTGAACCTGACATCAAACTTTCTGATATCGTGATACAATGGCTGAAGGAAGGTGTTTT
AGGCTTGGTCCATGAGTTCAAAGAAGGCAAAGATGAGCTGTGCGAGCAGGATGAAATGTT
CAGAGGCCGGACAGCAGTGTGCTGATCAAGTGATAGTTGGCAATGCCTCTTTGCGGCT
GAAAACGTGCAACTCACAGATGCTGGCACCTACAAATGTTATATCATCACTCTAAAGG
CAAGGGGAATGCTAACCTTGAGTATAAACTGGAGCCTCAGCATGCCGGAAGTCAATGT
30 GGACTATAATGCCAGCTCAGAGACCTTGCGGTGTGAGGCTCCCCGATGGTTCCCCAGCC
CACAGTGGTCTGGGCATCCCAAGTTGACCAGGGAGCCAACTTCTCGGAAGTCTCCAATAC
CAGCTTTGAGCTGAACCTCTGAGAATGTGACCATGAAGGTTGTGTCTGTGCTCTACAATGT
TACGATCAACAACACATACCTCTGTATGATTGAAAATGACATTGCCAAAGCAACAGGGGA
TATCAAAGTGACAGAATCGGAGATCAAAGGCGGAGTCACTACAGCTGCTAAACTCAA
35 GGCTTCTCTGTGTGTCTCTTCTTTCTTTGCCATCAGCTGGGCACCTTCTGCCCTCAGCCC
TTACCTGATGCTAAAATAATGTGCTCGGCCACAAAAAGCATGCAAAGTCATTGTTACA
ACAGGGATCTACAGAACTATTTACCACCAGATATGACCTAGTTTTATATTTCTGGGAGG
AAATGAATTCATATCTAGAAGTCTGGAGTGAGCAAAACAAGAGCAAGAAACAAAAAGAGC
CAAAGCAGAAGGCTCCAATATGAACAAGATAAATCTATCTTCAAAGACATATTAGAAGT
40 TGGGAAAATAATTCATGTGAAGTGAAGTCAACTGTGTGAGGGCTAAGAAACCTGGTTTT
GAGTAGAAAAGGGCTGGAAAGAGGGGAGCCAACAAATCTGTCTGCTTCCCTCACATTAGT
CATTTGGCAAATAAGCATTCTGTCTCTTTGGCTGCTGCCTCAGCACAGAGGCCAGAATC
TATCGGGCACCCAGGATAACATCTCTCAGTGAACAGAGTTGACAAGGCCATGGGAAATGC
CTGATGGGATTATCTTCAGCTTGTGAGCTTCTAAGTTTCTTTCCCTTCATTCTACCTG
45 CAAGCCAAAGTTCTGTAAGAGAAATGCCTGAGTTCTAGCTCAGGTTTTCTTACTCTGAATT
TAGATCTCCAGACCCTGCCTGGCCACAATTCAAATTAAGGCAACAAACATATACCTTCCA
TGAAGCACACAGACTTTTGAAGCAAGGACAATGACTGCTTGAATTGAGGCTTGAGG
AATGAAGCTTTGAAGGAAAAGATACTTTGTTTCCAGCCCCCTTCCACACTCTTCATGT
GTTAACCCTGCCTTCTGGACCTTGGAGCCACGGTGAAGTGTATTACATGTTGTTATAGA
50 AAAGTGAATTTAGAGTTCTGATCGTTCAAGAGAATGATTAAATATACATTTCTTAAAAAA
AAAAAAAAAAAA
>DEX0439_002.nt.2
TGTGAGTCAACAAGGAAGGCAGCGGCACTCCACTCAGCCAGTACCCAGATACGCTGGGAA
CCTTCCCCAGCCATGGCTTCCCTGGGGCAGATCCTCTTCTGGAGCATAATTAGCATCATC
ATTATTTCTGCTGGAGCAATTGCACTCATCATTGGCTTTGGTATTTAGAAAGTCTGTGTC
55 TGGCTTTTCAGCAATGAAGGGTTTGGTTGTAGAAGTTCCAAGGCTTCCCTTAGCATTGATC
TTTGCTTCCCTGAAGTGCAGGAGAGCACTCCATCAGTCACTACTGTGCGCTCAGCTGGG
AACATTGGGAGGATGGAATCCTGAGCTGCACTTTTGAACCTGACATCAAACTTTCTGAT
ATCGTGATACAATGGCTGAAGGAAGGTGTTTTAGGCTTGGTCCATGAGTTCAAAGAAGGC
AAAGATGAGCTGTGCGAGCAGGATGAAATGTTAGAGGCCGGACAGCAGTGTGCTGAT
60 CAAGTGATAGTTGGCAATGCCTCTTTGCGGCTGAAAACGTGCAACTCACAGATGCTGGC

ACCTACAAATGTTATATCATCACTTCTAAAGGCAAGGGGAATGCTAACCTTGAGTATAAA
ACTGGAGCCTTCAGCATGCCGGAAGTGAATGTGGACTATAATGCCAGCTCAGAGACCTTG
CGGTGTGAGGCTCCCCGATGGTTCAGCCACAGTGGTCTGGGCATCCCAAGTTGAC
CAGGGAGCCAACCTTCTCGGAAGTCTCCAATACCAGCTTGAGCTGAACCTTGAGAATGTG
5 ACCATGAAGGTTGTGTCTGTGCTCTACAATGTTACGATCAACAACATACTCCTGTATG
ATTGAAAAATGACATTGCCAAAGCAACAGGGGATATCAAAGTGACAGAATCGGAGATCAAA
AGGCGGAGTCACCTACAGCTGCTAAACTCAAAGGCTTCTCTGTGTGTCTCTTCTTTCTTT
GCCATCAGCTGGGCACTTCTGCCCTCTCAGCCCTTACCTGATGCTAAAATAATGTGCCTTG
10 GCCACAAAAAAGCATGCAAAGTCATTGTTACAACAGGGATCTACAGAATAATTCACCAC
CAGATATGACCTAGTTTTATATTTCTGGGAGGAAATGAATTCATATCTAGAAGTCTGGAG
TGAGCAACAAGAGCAAGAAACAAAAGAAGCCAAAAGCAGAAGGCTCCAATATGAACAA
GATAAATCTATCTTCAAAGACATATTAGAAGTTGGGAAAAATAATTCATGTGAAGTAGACA
AGTGTGTTAAGAGTGATAAGTAAATGCACGTGGAGACAAGTGCATCCCCAGATCTCAGG
GACCTCCCCCTGCCTGTACCTGGGGAGTGAGAGGACAGGATAGTGCATGTTCTTTGTCT
15 CTGAATTTTTAGTTATATGTGCTGTAATGTTGCTCTGAGGAAGCCCCCTGGAAAGTCTATC
CCAACATATCCACATCTTATATTCACAAATTAAGCTGTAGTATGTACCCTAAGACGCTG
CTAATTGACTGCCACTTCGCAACTCAGGGGCGGCTGCATTTTAGTAATGGGTCAAATGAT
TCACTTTTTATGATGCTTCCAAAGGTGCCTTGGCTTCTCTTCCCACTGACAAATGCCAA
AGTTGAGAAAAATGATCATAATTTTAGCATAAACAGAGCAGTCGCGCACACCGATTTTAT
20 AAATAAAGTGAACCTTCTTTTAAACAAACAAATGCGGGTTTATTTCTCAGATGATGT
TCATCCGTGAATGGTCCAGGGAAGGACCTTTACCTTGACTATATGGCATTATGTCATCA
CAAGCTCTGAGGCTTCTCTTTCCATCCTGCGTGGACAGCTAAGACCTCAGTTTTCAATA
GCATCTAGAGCAGTGGGACTCAGCTGGGGTGATTCGCCCCCATCTCCGGGGGAATGTC
TGAAGACAATTTTGGTTACCTCAATGAGGGAGTGGAGGAGGATACAGTGTACTACCAAC
25 TAGTGGATAAAGGCCAGGGATGCTGCTCAACCTCTACCATGTACAGGACGTCTCCCCAT
TACAACTACCCAATCCGAAGTGTCAACTGTGTGAGGACTAAGAAACCTGGTTTTGAGTA
GAAAAGGGCCTGGAAAGAGGGGAGCCAACAAATCTGTCTGCTTCTCATTAGTCATTG
GCAATAAAGCATTCTGTCTCTTTGGCTGCTGCTCAGCACAGAGAGCCAGAACTCTATCG
GGCACCAGGATAACATCTCTCAGTGAACAGAGTTGACAAGGCCTATGGGAAATGCCTGAT
30 GGGATTATCTTCAGCTTGTGAGCTTCTAAGTTTCTTCCCTTCATTCTACCCTGCAAGC
CAAGTTCTGTAAAGAGAAATGCCTGAGTTCTAGCTCAGGTTTTCTTACTCTGAATTTAGAT
CTCCAGACCTTCTTGGCCACAATTCAAATTAAGGCAACAAACATATACCTTCCATGAAG
CACACACAGACTTTTGAAGCAAGGACAATGACTGCTGAATTTAGGGCCTTGAGGAATGA
AGCTTTGAAGGAAAAGAATACTTTGTTTCCAGCCCCCTTCCACACTCTTCATGTGTTAA
35 CCACTGCCTTCTTGGACCTTGGAGCCACGGTGACTGTATTACATGTTGTTATAGAAAAT
GATTTTAGAGTTCTGATCGTTCAAGAGAATGATTAATATACATTTCTTA

>DEX0439_001.aa.1

MNRTWPRRIWGSSQDEAELIREDIQALHNYRSGRGERRAAALRATQEELQDRSPAAET
40 PPLQRRPSVRAVISTVERGAGRGRPOAKPIPEAEAAQRPEFVGTSSNADSASPDLGPRGP
DLAVLQAEREVDIINHVFDDES FVSRLQKSAEAAARVLEHRERGRSRRRAAGEGLLTLR
AKPPSEAEYTDVLQIKIYAFSLARLRGNIADPSSPELLHFLFGPLQMI VNTSGGPEFAS
SVRRPHLTSDAVALLRDNVT PRENELWTS LGDSWTRPGLELSPEEGPPYRPEFFSGWEPP
VTD PQSRAWEDPVEKQLQHERRRRQQSAPQVAVNGHRDLEPESEPOLESETAGKWVLCNY
45 DFQARNSSSELSVKQRDVLVLDLDSRKWWKVRDPAGQEGYVPYNI LTPYPGPRLHHSQSPA
RSLNSTPPPPPPAPAPPPALARPWRDRPRWDSCDSLNGLDPSEKEKFSQMLIVNEELQA
RLAQGRSGPSRAVPGPRAPEPQLSPGSDASEVRAWLQAKGFSSGTVDALGVLTGAQLFSL
QREELRAVSPEEGARVYSQVTVQRSLLDKEKVSELEAVMEKQKKKVEGEVEMEVI

>DEX0439_001.aa.2

MSPLSPGSFLPPLARADLTAILTGCPPLSACLVLAPRPHRRARLLPSEGLLTLRAKPPSE
50 AEYTDVLQIKIYAFSLARLRGNIADPSSPELLHFLFGPLQMI VNTSGGPEFASSVRPH
LTSDAVALLRDNVT PRENELWTS LGDSWTRPGLELSPEEGPPYRPEFFSGWEPPVTD PQS
RAWEDPVEKQLQHERRRRQQSAPQVAVNGHRDLEPESEPOLESETAGKWVLCNYDFQARN
SSELSVKQRDVLVLDLDSRKWWKVRDPAGQEGYVPYNI LTPYPGPRLHHSQSPARSLNST
55 PPPPPAPAPPPPPALARPWRDRPRWDSCDSLNGLDPSEKEKFSQMLIVNEELQARLAQGR
SGPSRAVPGPRAPEPQLSPGSDASEVRAWLQAKGFSSGTVDALGVLTGAQLFSLQKEELR
AVSPEEGARVYSQVTVQRSLLDKEKVSELEAVMEKQKKKVEGEVEMEVI

>DEX0439_001.aa.3

DLFQMSPLSPGSFLPPLARADLTAILTGCPPLSACLVLAPRPHRRARLLPSEGLLTLRAK
60 PPSEAEYTDVLQIKIYAFSLARLRGNIADPSSPELLHFLFGPLQMI VNTSGGPEFASSV

RRPHLTSDAVALLRDNVTPRENELWTSLGDSWTRPGLELSPEEGPPYRPEFFSGWEPPVT
DPQSRAWEDPVEKQLQHERRRRQQSAPQVAVNGHRDLEPESEPOLESETAGKWVLCNYDF
QARNSEL SVKQRDVLEVLDDSRKWWKVRDPAGQEGYVPYNILTPYPGPRLHHSQSPARS
5 LNSTFPFPPAPAPAPPALARPWRDRPRWDSCDSLNGLDPSEKEKFSQMLIVNEELQARL
AQGRSGPSRAVPGPRAPEPQLSPGSDASEVRAWLQAKGFSSGRRSCGR
>DEX0439_002.aa.1
MASLGQILFWISIISIIILAGAIALIIGFGISGRHSITVTTVASAGNIGEDGIQSCTFEP
DIKLSDIVIQWLKEGVLGLVHEFKEGKDELSEQDEMFRGRTAVFADQVIVGNASLRKKNV
10 QLTDA GTYKCYIITSKGKGNANLEYKTGAFSMPEVNVDYNASSETLRCEAPRWFPQPTVV
WASQVDQGANFSEVSNTSFELNSENVTMKVSVLYNVTINNTYSCMIENDIAKATGDIKV
TESEIKRRSHLQLLNSKASLCVSSFFAISWALLPLSPYIMLK
>DEX0439_002.aa.2
MASLGQILFWISIISIIILAGAIALIIGFGISEVSVWLSAMKGLVVEVPRLPALIFAS
>DEX0439_002.aa.3
15 TAGRHSITVTTVASAGNIGEDGILSCTFEPDIKLSDIVIQWLKEGVLGLVHEFKEGKDEL
SEQDEMFRGRTAVFADQVIVGNASLRKKNVQLTDAGTYKCYIITSKGKGNANLEYKTGAF
SMPEVNVDYNASSETLRCEAPRWFPQPTVVWASQVDQGANFSEVSNTSFELNSENVTMKV
VSVLYNVTINNTYSCMIENDIAKATGDIKVTESEIKRRSHLQLLNSKASLCVSSFFAISW
20 ALLPLSPYIMLK

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising:
 - (a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 10, 11, 13, 14;
 - (b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 2-6, 8;
 - (c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b); or
 - (d) a nucleic acid molecule having at least 95% sequence identity to the nucleic acid molecule of (a) or (b).
2. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a cDNA.
3. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is genomic DNA.
4. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is an RNA.
5. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a mammalian nucleic acid molecule.
6. The nucleic acid molecule according to claim 5, wherein the nucleic acid molecule is a human nucleic acid molecule.
7. A method for determining the presence of a ovarian specific nucleic acid (OSNA) in a sample, comprising the steps of:
 - (a) contacting the sample with the nucleic acid molecule of SEQ ID NO: 1-8 under conditions in which the nucleic acid molecule will selectively hybridize to a ovarian specific nucleic acid; and

(b) detecting hybridization of the nucleic acid molecule to a OSNA in the sample, wherein the detection of the hybridization indicates the presence of a OSNA in the sample.

- 5 8. A vector comprising the nucleic acid molecule of claim 1.
9. A host cell comprising the vector according to claim 8.
- 10 10. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of:
- 15 (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and
- (b) incubating the host cell under conditions in which the polypeptide is produced.
11. A polypeptide encoded by the nucleic acid molecule according to claim 1.
12. An isolated polypeptide selected from the group consisting of:
- 20 (a) a polypeptide comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 10, 11, 13, 14 ; or
- (b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 95% sequence identity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 2-6, 8.
- 25 13. An antibody or fragment thereof that specifically binds to:
- (a) a polypeptide comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 10, 11, 13, 14 ; or
- 30 (b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 95% sequence identity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 2-6, 8.
14. A method for determining the presence of a ovarian specific protein in a sample, comprising the steps of:

- (a) contacting the sample with a suitable reagent under conditions in which the reagent will selectively interact with the ovarian specific protein comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 9-14; and
- 5 (b) detecting the interaction of the reagent with a ovarian specific protein in the sample, wherein the detection of binding indicates the presence of a ovarian specific protein in the sample.
15. A method for diagnosing or monitoring the presence and metastases of ovarian cancer in a patient, comprising the steps of:
- 10 (a) determining an amount of:
- (i) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 9-14;
- (ii) a nucleic acid molecule comprising a nucleic acid sequence of SEQ
15 ID NO: 1-8;
- (iii) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (i) or (ii);
- (iv) a nucleic acid molecule having at least 95% sequence identity to the nucleic acid molecule of (i) or (ii);
- 20 (v) a polypeptide comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 9-14 ; or
- (vi) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 95% sequence identity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-8
25 and;
- (b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the ovarian specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the
30 presence of ovarian cancer.

16. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence of:

- (a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 9-14;
- 5 (b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-8;
- (c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b); or
- (d) a nucleic acid molecule having at least 95% sequence identity to the nucleic acid molecule of (a) or (b); or
- 10 (e) a polypeptide comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 9-14 ; or
- (f) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 95% sequence identity to a nucleic acid molecule
- 15 comprising a nucleic acid sequence of SEQ ID NO: 1-8.

17. A method of treating a patient with ovarian cancer, comprising the step of administering a composition consisting of:

- (a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 9-14;
- 20 (b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-8;
- (c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b);
- 25 (d) a nucleic acid molecule having at least 95% sequence identity to the nucleic acid molecule of (a) or (b);
- (e) a polypeptide comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 9-14 ; or
- (f) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 95% sequence identity to a nucleic acid molecule
- 30 comprising a nucleic acid sequence of SEQ ID NO: 1-8;

to a patient in need thereof, wherein said administration induces an immune response against the ovarian cancer cell expressing the nucleic acid molecule or polypeptide.

18. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 12.

ABSTRACT OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and polypeptides present in normal and neoplastic ovarian cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to

5 antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions containing the nucleic acid molecules, polypeptides, antibodies, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating ovarian cancer and non-cancerous

10 disease states in ovarian, identifying ovarian tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered ovarian tissue for treatment and research.

FIGURE 1

Ovr107	1	AGAGCAAGGAAGGGCAGGGGACCTGGGAAGGAAGTTCTGGAAGGCAGTGG	50
439_001.nt.2	1		0
Ovr107	51	GGTTTGAGATTGGACCCAGGGTCAAGATAGAACATGAAGGTGGGATGAGG	100
439_001.nt.2	1		0
Ovr107	101	ACATGAACAGAACATGGCCAAGAAGGATCTGGGGGAGCAGCCAGGACGAG	150
439_001.nt.2	1		0
Ovr107	151	GCGGAGCTGATCCGAGAGGACATCCAGGGGGCTCTGCACAATTACCGCTC	200
439_001.nt.2	1		0
Ovr107	201	GGGCCGCGGGGAGCGCAGGGCGGCGGCTCAGGGCCACGCAGGAGGAGT	250
439_001.nt.2	1		0
Ovr107	251	TGCAGCGCGACCGCTCGCCCGCCGCTGAGACCCCGCCCCTGCAGCGCCGC	300
439_001.nt.2	1		0
Ovr107	301	CCGTCAGTCCGCGCAGTGATCAGCACCGTAGAGCGGGGCGGGGCGCGGG	350
439_001.nt.2	1		0
Ovr107	351	ACGACCCCGAGCGAAGCCCATTCGCCAGGCAGAGGAGGCGCAGAGGCCTG	400
439_001.nt.2	1		0
Ovr107	401	AGCCGGTGGGGACCTCGAGCAA-----CGCTGACTC-----GGC-CTCC	438
439_001.nt.2	1	. . GATCTCTTCCAAATGTCCCCGCT--CTCCCCAGGCTCTCC	38
Ovr107	439	CCGGACCTGGGTCC-----CCGGGGTCCTGACCTGGCGGTTC TGCA	479
439_001.nt.2	39	. CC-----TCCCGCCACTTGCCAGGG--CTGACCT-----CA	67
Ovr107	480	-GGCGGAGCGGGAAGTGGACATCCTGAACCACGTGTTGACGACGTAGAG	528
439_001.nt.2	68 CCGC-----CAT-CTTAACCGGGTGTCC-----	89
Ovr107	529	AGCTTTGTATCGAGGCTGCAGAAGTCGGCGGAGGCGGCCAGGGTGCTGGA	578
439_001.nt.2	90 ACCTCTCT-----CTGC-----CTGCC-TGGTGCTGGC	116
Ovr107	579	GCACCGGGAACGCGGCCGCGAGGAGCCGGCGCCGGGCGGCT----GGGGAG	624
439_001.nt.2	117 CC-----CGCGTCCCA-TGCC-GCGCCCGTCTGCTCCCCTCAGAG	156

FIGURE 1 (continued)

Ovr107	625	GGCTTGCTGACGCTGCGGGCCAAGCCGCCCTCGGAGGCCGAGTACACCGA	674
439_001.nt.2	157	GGCTTGCTGACGCTGCGGGCCAAGCCGCCCTCGGAGGCCGAGTACACCGA	206
Ovr107	675	CGTGCTGCAGAAGATCAAGTACGCCTTCAGCCTGCTGGCCCGGCTGCGCG	724
439_001.nt.2	207	CGTGCTGCAGAAGATCAAGTACGCCTTCAGCCTGCTGGCCCGGCTGCGCG	256
Ovr107	725	GCAACATCGCCGACCCCTCCTCTCCGGAGCTGTTGCACTTCCTTTTCGGG	774
439_001.nt.2	257	GCAACATCGCCGACCCCTCCTCTCCGGAGCTGTTGCACTTCCTTTTCGGG	306
Ovr107	775	CCTCTGCAGATGATTGTGAACACGTCGGGGGGGCCGGAGTTCGCGAGCAG	824
439_001.nt.2	307	CCTCTGCAGATGATTGTGAACACGTCGGGGGGGCCGGAGTTCGCGAGCAG	356
Ovr107	825	TGTGCGGCGGCCGCATCTGACATCGGATGCCGTGGCGCTGCTGCGGGACA	874
439_001.nt.2	357	TGTGCGGCGGCCGCATCTGACATCGGATGCCGTGGCGCTGCTGCGGGACA	406
Ovr107	875	ACGTCACTCCACGTGAAAACGAGCTCTGGACCTCGCTGGGGGACTCGTGG	924
439_001.nt.2	407	ACGTCACTCCACGTGAAAACGAGCTCTGGACCTCGCTGGGGGACTCGTGG	456
Ovr107	925	ACCCGCCCCGGGCTGGAGCTGTCCCCGGAGGAGGGACCCCCATACAGACC	974
439_001.nt.2	457	ACCCGCCCCGGGCTGGAGCTGTCCCCGGAGGAGGGACCCCCATACAGACC	506
Ovr107	975	CGAGTTCCTTCAGCGGCTGGGAGCCGCCGGTCACTGACCCGCAGAGCCGCG	1024
439_001.nt.2	507	CGAGTTCCTTCAGCGGCTGGGAGCCGCCGGTCACTGACCCGCAGAGCCGCG	556
Ovr107	1025	CCTGGGAGGACCCAGTTGAGAAACAGCTACAGCACGAGCGGAGGCGCCGG	1074
439_001.nt.2	557	CCTGGGAGGACCCAGTTGAGAAACAGCTACAGCACGAGCGGAGGCGCCGG	606
Ovr107	1075	CAGCAAAGCGCCCCCAGGTCGCTGTCAATGGTCACCGAGACTTGGAGCC	1124
439_001.nt.2	607	CAGCAAAGCGCCCCCAGGTCGCTGTCAATGGTCACCGAGACTTGGAGCC	656
Ovr107	1125	AGAATCTGAGCCTCAGCTGGAGTCAGAGACAGCAGGAAAATGGGTCCTGT	1174
439_001.nt.2	657	AGAATCTGAGCCTCAGCTGGAGTCAGAGACAGCAGGAAAATGGGTCCTGT	706
Ovr107	1175	GTAATTATGACTTCCAGGCCCGCAACAGCAGTGAGCTGTTCGGTCAAGCAG	1224
439_001.nt.2	707	GTAATTATGACTTCCAGGCCCGCAACAGCAGTGAGCTGTTCGGTCAAGCAG	756
Ovr107	1225	CGGGACGTACTGGAGGTCCTGGATGACAGTCGTAAGTGGTGGAAAGGTTCC	1274
439_001.nt.2	757	CGGGACGTACTGGAGGTCCTGGATGACAGTCGTAAGTGGTGGAAAGGTTCC	806

FIGURE 1 (continued)

Ovr107	1275	GGACCCAGCGGGGCAGGAGGGATATGTGCCCTACAACATCCTGACACCCCT	1324
439_001.nt.2	807		856
Ovr107	1325	ACCCCGGACCCCGGCTGCACCACAGCCAAAGCCCTGCCCCGAGCCTGAAC	1374
439_001.nt.2	857		906
Ovr107	1375	AGCACTCCTCCTCCACCACCAGCCCCAGCCCCGGCCCCACCTCCAGCTCT	1424
439_001.nt.2	907		956
Ovr107	1425	GGCTCGGCCCCGCTGGGACAGGCCCCGCTGGGACAGCTGCGATAGCCTCA	1474
439_001.nt.2	957		1006
Ovr107	1475	ACGGCTTGGACCCAGCGAGAAGGAGAAATTCTCCAGATGCTCATCGTC	1524
439_001.nt.2	1007		1056
Ovr107	1525	AACGAGGAACTGCAGGCGCGCTTGGCCCAGGGCCGCTCGGGACCGAGCCG	1574
439_001.nt.2	1057		1106
Ovr107	1575	CGCAGTCCCAGGGCCCCGCGCCCCGGAACCGCAGCTCAGCCCCGGGCTCGG	1624
439_001.nt.2	1107		1156
Ovr107	1625	ACGCCTCCGAGGTCCGCGCCTGGCTGCAGGCCAAGGGCTTTAGCTCCGGG	1674
439_001.nt.2	1157		1204
Ovr107	1675	ACCGTGGACGCGCTGGGTGTGCTGACCGGGGCGCAGCTTTTCTCGCTGCA	1724
439_001.nt.2	1205	-----	1204
Ovr107	1725	GAGGGAGGAGCTGCGGGCGGTGAGCCCCGAGGAGGGGGCACGTGTGTACA	1774
439_001.nt.2	1205	.	1254
Ovr107	1775	GCCAGGTCACCGTGACGCGCTCGCTGCTGGAGGACAAAGAGAAAGTGTCA	1824
439_001.nt.2	1255		1304
Ovr107	1825	GAGCTGGAGGCAGTGATGGAGAAGCAAAAGAAGAAGGTGGAAGGCGAGGT	1874
439_001.nt.2	1305		1354
Ovr107	1875	GGAAATGGAGGTCATTTGACCTGCCAGGCGCCCTTCGCAAAGAGTGACGA	1924
439_001.nt.2	1355		1404

FIGURE 1 (continued)

Ovr107	1925	GGCCCCGTGGGAGAACGGACTCCTCAGACTCTCCCCAATAGCGGAAGTCG	1974
439_001.nt.2	1405	GGCCCCGTGGGAGAACGGACTCCTCAGACTCTCCCCAATAGCGGAAGTCG	1454
Ovr107	1975	ATCTTCTGAAGGATGGCCAATCTGCTCCGGCCCTGGTCTTCCCCCATCCC	2024
439_001.nt.2	1455	ATCTTCTGAAGGATGGCCAATCTGCTCCGGCCCTGGTCTTCCCCCATCCC	1504
Ovr107	2025	GGTGGACAGACTTAACGATCCTTGCTGCAGTCCCTCCGGAGAGGATCTGG	2074
439_001.nt.2	1505	GGTGGACAGACTTAACGATCCTTGCTGCAGTCCCTCCGGAGAGGATCTGG	1554
Ovr107	2075	ACTGGCTGGGAGTGGGGAGGGCGTGGAGACAGTCTACGGAAAGCGCTAGC	2124
439_001.nt.2	1555	ACTGGCTGGGAGTGGGGAGGGCGTGGAGACAGTCTACGGAAAGCGCTAGC	1604
Ovr107	2125	AGACCCCCGAGAGGGTGCAGTGGAGCCCTGAGCATTGTAATATGCGGCCC	2174
439_001.nt.2	1605	AGACCCCCGAGAGGGTGCAGTGGAGCCCTGAGCATTGTAATATGCGGCCC	1654
Ovr107	2175	AGCCTATAAACAGCCTCCGTGCTTAGCAAAAAAAAAAAAAAAAAAAAAA	2221
439_001.nt.2	1655	AGCCTATAAACAGCCTCCGTGCTTAGCAG	1683

FIGURE 2

Ovr107_aa	1	MNRTWPRRIWGSSQDEAEILIREDIQGA LHNYRSGRGERRAAALRATQEEI	50
439_001.aa.2	1		0
Ovr107_aa	51	QDRDSPAAETPPLQRRPSVRAVISTVERGAGRGRPQAKPIPEAEAAQORPE	100
439_001.aa.2	1		0
Ovr107_aa	101	PVGTSSNADSASPDLGPRGPD LAVLQAEREVDILNHVFDDVESFVSRLQK	150
439_001.aa.2	1	MSP-LSPGSP LPP LARAD-----LTAILTG	24
Ovr107_aa	151	SAEAA RVLEHREGRRSRRRAAGEGLLTLRAKPPSEAEYTDVLQIKIYAF	200
439_001.aa.2	25	CPPLSACLVLAPRP HRRARLLPSEGLLTLRAKPPSEAEYTDVLQIKIYAF	74
Ovr107_aa	201	SLLARLRGNIADPSSPELLHFLFGP LQMI VNTSGGPEF ASSVRPHLTS D	250
439_001.aa.2	75	SLLARLRGNIADPSSPELLHFLFGP LQMI VNTSGGPEF ASSVRPHLTS D	124
Ovr107_aa	251	AVALLRDNVTPRENELWTS LGDSWTRPGLELSPEEGPPYRPEFFSGWEPP	300
439_001.aa.2	125	AVALLRDNVTPRENELWTS LGDSWTRPGLELSPEEGPPYRPEFFSGWEPP	174
Ovr107_aa	301	VTDPQSRAWEDFVEKQLQHERRRRQQSAPQVAVNGHRDLEPESEFQLESE	350
439_001.aa.2	175	VTDPQSRAWEDFVEKQLQHERRRRQQSAPQVAVNGHRDLEPESEFQLESE	224
Ovr107_aa	351	TAGKWVLCNYDFQARNSSSELSVKQ RDVLEVLDDSRKWWKV RDPAGQEGYV	400
439_001.aa.2	225	TAGKWVLCNYDFQARNSSSELSVKQ RDVLEVLDDSRKWWKV RDPAGQEGYV	274
Ovr107_aa	401	PYNILTPYPG PRLHHSQSPARSLNSTPPPPPPAPAPAPPPALARPWRDRPR	450
439_001.aa.2	275	PYNILTPYPG PRLHHSQSPARSLNSTPPPPPPAPAPAPPPALARPWRDRPR	324
Ovr107_aa	451	WDSCDSLNGLD PSEKEKFSQMLIVNEELQARLAQGRSGPSRAVPGPRAPE	500
439_001.aa.2	325	WDSCDSLNGLD PSEKEKFSQMLIVNEELQARLAQGRSGPSRAVPGPRAPE	374
Ovr107_aa	501	PQLSPGSDASEVRAWLQAKGFSSGTVDALGVLTGAQLFS LQREELRAVSP	550
439_001.aa.2	375	PQLSPGSDASEVRAWLQAKGFSSGTVDALGVLTGAQLFS LQKEELRAVSP	424
Ovr107_aa	551	EEGARVYSQVTVQ RSLLEDKEKVSELEAVMEKQKKKVEGEVEMEVI	596
439_001.aa.2	425	EEGARVYSQVTVQ RSLLEDKEKVSELEAVMEKQKKKVEGEVEMEVI	470

FIGURE 3

Ovr107_aa	1	MNRTWPRRIWGSSQDEAELIREDIQGA LHNYRSGRGERRAAALRATQEEL	50
439_001.aa.3	1		0
Ovr107_aa	51	QRDRSPAETPPLQRRPSVRAVISTVERGAGRGRFPQAKPIPEAEAAQRPE	100
439_001.aa.3	1		0
Ovr107_aa	101	FVGTSSNADSASPD LGPRGFDLAVLQAEREVDI LNHVFD DVESFVSRLQK	150
439_001.aa.3	1	DLFQMSP-LSPGSP LPLARAD-----LTAILTG	28
Ovr107_aa	151	SAEAA RVLEHRRERGRSRRRAAGEGLLTLRAKPPSEAEYTDVLQIKIYAF	200
439_001.aa.3	29	CPPLSACLVLAPRPHRRARLLPSEGLLTLRAKPPSEAEYTDVLQIKIYAF	78
Ovr107_aa	201	SLLARLRGNIADPSSPELLHFLFGPLQMI VNTSGGPEFASSVRRPHLTSD	250
439_001.aa.3	79	SLLARLRGNIADPSSPELLHFLFGPLQMI VNTSGGPEFASSVRRPHLTSD	128
Ovr107_aa	251	AVALLRDNVTPRENELWTS LGDSWTRPGLELSPEEGPPYRPEFFSGWEPP	300
439_001.aa.3	129	AVALLRDNVTPRENELWTS LGDSWTRPGLELSPEEGPPYRPEFFSGWEPP	178
Ovr107_aa	301	VTDPQSRRAWEDFVEKQLQHERRRRQQSAPQVAVNGHRDLEPESEPOLESE	350
439_001.aa.3	179	VTDPQSRRAWEDFVEKQLQHERRRRQQSAPQVAVNGHRDLEPESEPOLESE	228
Ovr107_aa	351	TAGKWVLCNYDFQARNSS ELSVKQRDVLEVLDDSRKWWKVRDPAGQEGYV	400
439_001.aa.3	229	TAGKWVLCNYDFQARNSS ELSVKQRDVLEVLDDSRKWWKVRDPAGQEGYV	278
Ovr107_aa	401	PYNILTPYPGPR LHHSQSPARSLNSTPPPPPPAPAPAPFPFALARPWRDRPR	450
439_001.aa.3	279	PYNILTPYPGPR LHHSQSPARSLNSTPPPPPPAPAPAPFPFALARPWRDRPR	328
Ovr107_aa	451	WDSCDSINGLDPSEKEKFSQMLIVNEELQARLAQGRSGPSRAVFGPRAPE	500
439_001.aa.3	329	WDSCDSINGLDPSEKEKFSQMLIVNEELQARLAQGRSGPSRAVFGPRAPE	378
Ovr107_aa	501	PQLSPGSDASEVRAWLQAKGFSSGTV DALGVLTGAQLFSIQREELRAVSP	550
439_001.aa.3	379	PQLSPGSDASEVRAWLQAKGFSSGRRSCGR	408
Ovr107_aa	551	EEGARVYSQVTVQRSLL EDKEKVSELEAVMEKQKKKVEGEVEMEVI	596
439_001.aa.3	409		408

FIGURE 4

Ovr107	1	AGAGCAAGGAAGGGCAGGGGACCTGGGAAGGAAGTTCTGGAAGGCAGTGG	50
439_001.nt.4	1		0
Ovr107	51	GGTTTGAGATTGGACCCAGGGTCAAGATAGAACATGAAGGTGGGATGAGG	100
439_001.nt.4	1		0
Ovr107	101	ACATGAACAGAACATGGCCAAGAAGGATCTGGGGGAGCAGCCAGGACGAG	150
439_001.nt.4	1		0
Ovr107	151	GCGGAGCTGATCCGAGAGGACATCCAGGGGGCTCTGCACAATTACCGCTC	200
439_001.nt.4	1		0
Ovr107	201	GGGCCGCGGGGAGCGCAGGGCGGGCGGCTCAGGGCCACGCAGGAGGAGT	250
439_001.nt.4	1		0
Ovr107	251	TGCAGCGCGACCGCTCGCCCGCCGCTGAGACCCCGCCCCCTGCAGCGCCGC	300
439_001.nt.4	1		0
Ovr107	301	CCGTCAGTCCGCGCAGTGATCAGCACCGTAGAGCGGGGCGCGGGCCGCGG	350
439_001.nt.4	1		0
Ovr107	351	ACGACCCCAGGCGAAGCCCATTCCTCGAGGCAGAGGAGGCGCAGAGGCCTG	400
439_001.nt.4	1		0
Ovr107	401	AGCCGGTGGGGACCTCGAGCAA-----CGCTGACTC----GGC-CTCC	438
439_001.nt.4	1 GATCTCTTCCAAATGTCCCGCT--CTCCCCAGGCTCTCC	38
Ovr107	439	CCGGACCTGGGTCC-----CCGGGGTCCTGACCTGGCGGTTCTGCA	479
439_001.nt.4	39	. CC-----TCCCGCCACTTGCCAGGG--CTGACCT-----CA	67
Ovr107	480	-GGCGGAGCGGGAAGTGGACATCCTGAACCACGTGTTTCGACGACGTAGAG	528
439_001.nt.4	68 CCGC-----CAT-CTTAACGGGTGTCC-----	89
Ovr107	529	AGCTTTGTATCGAGGCTGCAGAAAGTCGGCGGAGGCGGCCAGGGTGCTGGA	578
439_001.nt.4	90 ACCTCTCT-----CTGC-----CTGCC-TGGTGCTGGC	116
Ovr107	579	GCACCGGGAACGCGGCCGAGGAGCCGGCGCCGGGCGGGCT----GGGGAG	624
439_001.nt.4	117 CC-----CGCGTCCCCA-TGCC-GCGCCGCTCTGCTCCCTCAGAG	156

FIGURE 4 (continued)

Ovr107	625	GGCTTGCTGACGCTGCGGGCCAAGCCGCCCTCGGAGGCCGAGTACACCGA	674
439_001.nt.4	157	GGCTTGCTGACGCTGCGGGCCAAGCCGCCCTCGGAGGCCGAGTACACCGA	206
Ovr107	675	CGTGCTGCAGAAGATCAAGTACGCCTTCAGCCTGCTGGCCCGGCTGCGCG	724
439_001.nt.4	207	CGTGCTGCAGAAGATCAAGTACGCCTTCAGCCTGCTGGCCCGGCTGCGCG	256
Ovr107	725	GCAACATCGCCGACCCCTCCTCTCCGGAGCTGTTGCACTTCCTTTTCGGG	774
439_001.nt.4	257	GCAACATCGCCGACCCCTCCTCTCCGGAGCTGTTGCACTTCCTTTTCGGG	306
Ovr107	775	CCTCTGCAGATGATTGTGAACACGTCGGGGGGGCCGGAGTTCGCGAGCAG	824
439_001.nt.4	307	CCTCTGCAGATGATTGTGAACACGTCGGGGGGGCCGGAGTTCGCGAGCAG	356
Ovr107	825	TGTGCGGCGGCCCGCATCTGACATCGGATGCCGTGGCGCTGCTGCGGGACA	874
439_001.nt.4	357	TGTGCGGCGGCCCGCATCTGACATCGGATGCCGTGGCGCTGCTGCGGGACA	406
Ovr107	875	ACGTCACTCCACGTGAAAACGAGCTCTGGACCTCGCTGGGGGACTCGTGG	924
439_001.nt.4	407	ACGTCACTCCACGTGAAAACGAGCTCTGGACCTCGCTGGGGGACTCGTGG	456
Ovr107	925	ACCCGCCCCGGGCTGGAGCTGTCCCCGGAGGAGGGACCCCCATACAGACC	974
439_001.nt.4	457	ACCCGCCCCGGGCTGGAGCTGTCCCCGGAGGAGGGACCCCCATACAGACC	506
Ovr107	975	CGAGTTCTTCAGCGGCTGGGAGCCGCCGGTCAC TGACCCGCAGAGCCGCG	1024
439_001.nt.4	507	CGAGTTCTTCAGCGGCTGGGAGCCGCCGGTCAC TGACCCGCAGAGCCGCG	556
Ovr107	1025	CCTGGGAGGACCCAGTTGAGAAACAGCTACAGCACGAGCGGAGGCGCCGG	1074
439_001.nt.4	557	CCTGGGAGGACCCAGTTGAGAAACAGCTACAGCACGAGCGGAGGCGCCGG	606
Ovr107	1075	CAGCAAAGCGCCCCCAGGTCGCTGTCAATGGTCACCGAGACTTGGAGCC	1124
439_001.nt.4	607	CAGCAAAGCGCCCCCAGGTCGCTGTCAATGGTCACCGAGACTTGGAGCC	656
Ovr107	1125	AGAATCTGAGCCTCAGCTGGAGTCAGAGACAGCAGGAAAATGGGTCCTGT	1174
439_001.nt.4	657	AGAATCTGAGCCTCAGCTGGAGTCAGAGACAGCAGGAAAATGGGTCCTGT	706
Ovr107	1175	GTAATTATGACTTCCAGGCCCGCAACAGCAGTGAGCTGTCCGTCAAGCAG	1224
439_001.nt.4	707	GTAATTATGACTTCCAGGCCCGCAACAGCAGTGAGCTGTCCGTCAAGCAG	756
Ovr107	1225	CGGGACGTACTGGAGGTCCTGGATGACAGTCGTAAGTGGTGGAAAGGTTCC	1274
439_001.nt.4	757	CGGGACGTACTGGAGGTCCTGGATGACAGTCGTAAGTGGTGGAAAGGTTCC	806

FIGURE 4 (continued)

Ovr107	1275	GGACCCAGCGGGGAGGAGGGATATGTGCCCTACAACATCCTGACACCCT	1324
439_001.nt.4	807	GGACCCAGCGGGGAGGAGGGATATGTGCCCTACAACATCCTGACACCCT	856
Ovr107	1325	ACCCCGGACCCCGGCTGCACCACAGCCAAAGCCCTGCCCGCAGCCTGAAC	1374
439_001.nt.4	857	ACCCCGGACCCCGGCTGCACCACAGCCAAAGCCCTGCCCGCAGCCTGAAC	906
Ovr107	1375	AGCACTCCTCCTCCACCACCAGCCCCAGCCCCGGCCCCACCTCCAGCTCT	1424
439_001.nt.4	907	AGCACTCCTCCTCCACCACCAGCCCCAGCCCCGGCCCCACCTCCAGCTCT	956
Ovr107	1425	GGCTCGGCCCCGCTGGGACAGGCCCCGCTGGGACAGCTGCGATAGCCTCA	1474
439_001.nt.4	957	GGCTCGGCCCCGCTGGGACAGGCCCCGCTGGGACAGCTGCGATAGCCTCA	1006
Ovr107	1475	ACGGCTTGGACCCCAGCGAGAAGGAGAAATTCTCCAGATGCTCATCGTC	1524
439_001.nt.4	1007	ACGGCTTGGACCCCAGCGAGAAGGAGAAATTCTCCAGATGCTCATCGTC	1056
Ovr107	1525	AACGAGGAAC TGAGGCGCGCTGGCCCAGGGCCGCTCGGGACCGAGCCG	1574
439_001.nt.4	1057	AACGAGGAAC TGAGGCGCGCTGGCCCAGGGCCGCTCGGGACCGAGCCG	1106
Ovr107	1575	CGCAGTCCCAGGGCCCCGCGCCCCGGAACCGCAGCTCAGCCCCGGGCTCGG	1624
439_001.nt.4	1107	CGCAGTCCCAGGGCCCCGCGCCCCGGAACCGCAGCTCAGCCCCGGGCTCGG	1156
Ovr107	1625	ACGCCTCCGAGGTCCGCGCCTGGCTGCAGGCCAAGGGCTTTAGCTCCGGG	1674
439_001.nt.4	1157	ACGCCTCCGAGGTCCGCGCCTGGCTGCAGGCCAAGGGCTTTAGCTCCGGG	1206
Ovr107	1675	ACCGTGGACGCGCTGGGTGTGCTGACCGGGGCGCAGCTTTTCTCGCTGCA	1724
439_001.nt.4	1207	ACCGTGGACGCGCTGGGTGTGCTGACCGGGGCGCAGCTTTTCTCGCTGCA	1256
Ovr107	1725	GAGGGAGGAGCTGCGGGCGGTGAGCCCCGAGGAGGGGGCACGTGTGTACA	1774
439_001.nt.4	1257	GAAGGAGGAGCTGCGGGCGGTGAGCCCCGAGGAGGGGGCACGTGTGTACA	1306
Ovr107	1775	GCCAGGTCACCGTGACGCGCTCGCTGCTGGAGGACAAAGAGAAAGTGTCA	1824
439_001.nt.4	1307	GCCAGGTCACCGTGACGCGCTCGCTGCTGGAGGACAAAGAGAAAGTGTCA	1356
Ovr107	1825	GAGCTGGAGGCAGTGATGGAGAAGCAAAAGAAGAAGGTGGAAGGCGAGGT	1874
439_001.nt.4	1357	GAGCTGGAGGCAGTGATGGAGAAGCAAAAGAAGAAGGTGGAAGGCGAGGT	1406
Ovr107	1875	GGAAATGGAGGTCATTTGACCTGCCAGGCGCCCTTCGCAAAGAGTGACGA	1924
439_001.nt.4	1407	GGAAATGGAGGTCATTTGACCTGCCAGGCGCCCTTCGCAAAGAGTGACGA	1456

FIGURE 4 (continued)

Ovr107	1925	GGCCCCGTGGGAGAACGGACTCCTCAGACTCTCCCCAATAGCGGAAGTCG	1974
439_001.nt.4	1457	GGCCCCGTGGGAGAACGGACTCCTCAGACTCTCCCCAATAGCGGAAGTCG	1506
Ovr107	1975	ATCTTCTGAAGGATGGCCAATCTGCTCCGGCCCTGGTCTTCCCCCATCCC	2024
439_001.nt.4	1507	ATCTTCTGAAGGATGGCCAATCTGCTCCGGCCCTGGTCTTCCCCCATCCC	1556
Ovr107	2025	GGTGGACAGACTTAACGATCCTTGCTGCAGTCCCTCCGGAGAGGATCTGG	2074
439_001.nt.4	1557	GGTGGACAGACTTAACGATCCTTGCTGCAGTCCCTCCGGAGAGGATCTGG	1606
Ovr107	2075	ACTGGCTGGGAGTGGGGAGGGCGTGGAGACAGTCTACGGAAAGCGCTAGC	2124
439_001.nt.4	1607	ACTGGCTGGGAGTGGGGAGGGCGTGGAGACAGTCTACGGAAAGCGCTAGC	1656
Ovr107	2125	AGACCCCCGAGAGGGTGCAGTGGAGCCCTGAGCATTGTAATATGCGGCCC	2174
439_001.nt.4	1657	AGACCCCCGAGAGGGTGCAGTGGAGCCCTGAGCATTGTAATATGCGGCCC	1706
Ovr107	2175	AGCCTATAAACAGCCTCCGTGCTTAGCAAAAAAAAAAAAAAAAAAAAAA	2221
439_001.nt.4	1707	AGCCTATAAACAGCCTCCGTGCTTAGCAG	1735

FIGURE 5

Ovr107	1	AGAGCAAGGAAGGGCAGGGGACCTGGGAAGGAAGTTCTGGAAGGCAGTGG	50
439_001.nt.6	1		0
Ovr107	51	GGTTTGAGATTGGACCCAGGGTCAAGATAGAACATGAAGGTGGGATGAGG	100
439_001.nt.6	1		0
Ovr107	101	ACATGAACAGAACATGGCCAAGAAGGATCTGGGGGAGCAGCCAGGACGAG	150
439_001.nt.6	1		0
Ovr107	151	GCGGAGCTGATCCGAGAGGACATCCAGGGGGCTCTGCACAATTACCGCTC	200
439_001.nt.6	1		0
Ovr107	201	GGGCCGCGGGGAGCGCAGGGCGGGCGGCTCAGGGCCACGCAGGAGGAGT	250
439_001.nt.6	1		0
Ovr107	251	TGCAGCGCGACCGCTCGCCCGCCGCTGAGACCCCGCCCCTGCAGCGCCGC	300
439_001.nt.6	1		0
Ovr107	301	CCGTCAGTCCGCGCAGTGATCAGCACCGTAGAGCGGGGCGCGGGCCGCGG	350
439_001.nt.6	1		0
Ovr107	351	ACGACCCAGGCGAAGCCCATTTCCCGAGGCAGAGGAGGCGCAGAGGCCTG	400
439_001.nt.6	1		0
Ovr107	401	AGCCGGTGGGGACCTCGAGCAA-----CGCTGACTC----GGC-CTCC	438
439_001.nt.6	1	. . GATCTCTTCCAAATGTCCCGCT--CTCCCCAGGCTCTCC	38
Ovr107	439	CCGGACCTGGGTCC-----CCGGGGTCCTGACCTGGCGGTTCTGCA	479
439_001.nt.6	39	. CC-----TCCCGCCACTTGCCAGGG--CTGACCT-----CA	67
Ovr107	480	-GGCGGAGCGGGAAGTGGACATCCTGAACCACGTGTTGACGACGTAGAG	528
439_001.nt.6	68 CCGC-----CAT-CTTAACCGGGTGTCC-----	89
Ovr107	529	AGCTTTGTATCGAGGCTGCAGAAGTCGGCGGAGGCGGCCAGGGTGCTGGA	578
439_001.nt.6	90 ACCTCTCT-----CTGC-----CTGCC-TGGTGCTGGC	116
Ovr107	579	GCACCGGGAACGCGGCCGAGGAGCCGGCGCGGGCGGCT----GGGGAG	624
439_001.nt.6	117 CC-----CGCGTCCCCA-TCGCC-GCGCCCGTCTGCTCCCCCTCAGAG	156

FIGURE 5 (continued)

Ovr107	625	GGCTTGCTGACGCTGCGGGCCAAGCCGCCCTCGGAGGCCGAGTACACCGA	674
439_001.nt.6	157	GGCTTGCTGACGCTGCGGGCCAAGCCGCCCTCGGAGGCCGAGTACACCGA	206
Ovr107	675	CGTGCTGCAGAAGATCAAGTACGCCTTCAGCCTGCTGGCCCCGGCTGCGCG	724
439_001.nt.6	207	CGTGCTGCAGAAGATCAAGTACGCCTTCAGCCTGCTGGCCCCGGCTGCGCG	256
Ovr107	725	GCAACATCGCCGACCCCTCCTCTCCGGAGCTGTTGCAC TTCCTTTTCGGG	774
439_001.nt.6	257	GCAACATCGCCGACCCCTCCTCTCCGGAGCTGTTGCAC TTCCTTTTCGGG	306
Ovr107	775	CCTCTGCAGATGATTGTGAACACGTCGGGGGGGCCGGAGTTCGCGAGCAG	824
439_001.nt.6	307	CCTCTGCAGATGATTGTGAACACGTCGGGGGGGCCGGAGTTCGCGAGCAG	356
Ovr107	825	TGTGCGGCGGCCGCATCTGACATCGGATGCCGTGGCGCTGCTGCGGGACA	874
439_001.nt.6	357	TGTGCGGCGGCCGCATCTGACATCGGATGCCGTGGCGCTGCTGCGGGACA	406
Ovr107	875	ACGTCACCTCCACGTGAAAACGAGCTCTGGACCTCGCTGGGGGACTCGTGG	924
439_001.nt.6	407	ACGTCACCTCCACGTGAAAACGAGCTCTGGACCTCGCTGGGGGACTCGTGG	456
Ovr107	925	ACCCGCCCCGGGCTGGAGCTGTCCCCGGAGGAGGGACCCCCATACAGACC	974
439_001.nt.6	457	ACCCGCCCCGGGCTGGAGCTGTCCCCGGAGGAGGGACCCCCATACAGACC	506
Ovr107	975	CGAGTTCTTCAGCGGCTGGGAGCCGCCGGTCACTGACCCGCAGAGCCGCG	1024
439_001.nt.6	507	CGAGTTCTTCAGCGGCTGGGAGCCGCCGGTCACTGACCCGCAGAGCCGCG	556
Ovr107	1025	CCTGGGAGGACCCAGTTGAGAAACAGCTACAGCACGAGCGGAGGCGCCGG	1074
439_001.nt.6	557	CCTGGGAGGACCCAGTTGAGAAACAGCTACAGCACGAGCGGAGGCGCCGG	606
Ovr107	1075	CAGCAAAGCGCCCCCAGGTCGCTGTCAATGGTCACCGAGACTTGGAGCC	1124
439_001.nt.6	607	CAGCAAAGCGCCCCCAGGTCGCTGTCAATGGTCACCGAGACTTGGAGCC	656
Ovr107	1125	AGAATCTGAGCCTCAGCTGGAGTCAGAGACAGCAGGAAAATGGGTCTCTGT	1174
439_001.nt.6	657	AGAATCTGAGCCTCAGCTGGAGTCAGAGACAGCAGGAAAATGGGTCTCTGT	706
Ovr107	1175	GTAATTATGACTTCCAGGCCCGCAACAGCAGTGAGCTGTCCGGTCAAGCAG	1224
439_001.nt.6	707	GTAATTATGACTTCCAGGCCCGCAACAGCAGTGAGCTGTCCGGTCAAGCAG	756
Ovr107	1225	CGGGACGTACTGGAGGTCCTGGATGACAGTCGTAAGTGGTGGAAGGTTTCG	1274
439_001.nt.6	757	CGGGACGTACTGGAGGTCCTGGATGACAGTCGTAAGTGGTGGAAGGTTTCG	806

FIGURE 5 (continued)

Ovr107	1275	GGACCCAGCGGGGCAGGAGGGATATGTGCCCTACAACATCCTGACACCCT	1324
439_001.nt.6	807	GGACCCAGCGGGGCAGGAGGGATATGTGCCCTACAACATCCTGACACCCT	856
Ovr107	1325	ACCCCGGACCCCGGCTGCACCACAGCCAAAGCCCTGCCCGCAGCCTGAAC	1374
439_001.nt.6	857	ACCCCGGACCCCGGCTGCACCACAGCCAAAGCCCTGCCCGCAGCCTGAAC	906
Ovr107	1375	AGCACTCCTCCTCCACCACCAGCCCCAGCCCCGGCCCCACCTCCAGCTCT	1424
439_001.nt.6	907	AGCACTCCTCCTCCACCACCAGCCCCAGCCCCGGCCCCACCTCCAGCTCT	956
Ovr107	1425	GGCTCGGCCCCGCTGGGACAGGCCCCGCTGGGACAGCTGCGATAGCCTCA	1474
439_001.nt.6	957	GGCTCGGCCCCGCTGGGACAGGCCCCGCTGGGACAGCTGCGATAGCCTCA	1006
Ovr107	1475	ACGGCTTGGACCCCAGCGAGAAGGAGAAATTCTCCAGATGCTCATCGTC	1524
439_001.nt.6	1007	ACGGCTTGGACCCCAGCGAGAAGGAGAAATTCTCCAGATGCTCATCGTC	1056
Ovr107	1525	AACGAGGAACTGCAGGCGCGCCTGGCCCAGGGCCGCTCGGGACCGAGCCG	1574
439_001.nt.6	1057	AACGAGGAACTGCAGGCGCGCCTGGCCCAGGGCCGCTCGGGACCGAGCCG	1106
Ovr107	1575	CGCAGTCCCAGGGCCCCGCGCCCCGGAACCGCAGCTCAGCCCCGGGCTCGG	1624
439_001.nt.6	1107	CGCAGTCCCAGGGCCCCGCGCCCCGGAACCGCAGCTCAGCCCCGGGCTCGG	1156
Ovr107	1625	ACGCCTCCGAGGTCCGCGCCTGGCTGCAGGCCAAGGGCTTTAGCTCCGGG	1674
439_001.nt.6	1157	ACGCCTCCGAGGTCCGCGCCTGGCTGCAGGCCAAGGGCTTTAGCTCCGGG	1206
Ovr107	1675	ACCGTGGACGCGCTGGGTGTGCTGACCGGGGCGCAGCTTTTCTCGCTGCA	1724
439_001.nt.6	1207	ACCGTGGACGCGCTGGGTGTGCTGACCGGGGCGCAGCTTTTCTCGCTGCA	1256
Ovr107	1725	GAGGGAGGAGCTGCGGGCGGTGAGCCCCGAGGAGGGGGCACGTGTGTACA	1774
439_001.nt.6	1257	GAAGGAGGAGCTGCGGGCGGTGAGCCCCGAGGAGGGGGCACGTGTGTACA	1306
Ovr107	1775	GCCAGGTCACCGTGCAGCGCTCGCTGCTGGAGGACAAAGAGAAAGTGTCA	1824
439_001.nt.6	1307	GCCAGGTCACCGTGCAGCGCTCGCTGCTGGAGGACAAAGAGAAAGTGTCA	1356
Ovr107	1825	GAGCTGGAGGCAGTGATGGAGAAGCAAAAGAAGAAGGTGGAAGGCGAGGT	1874
439_001.nt.6	1357	GAGCTGGAGGCAGTGATGGAGAAGCAAAAGAAGAAGGTGGAAGGCGAGGT	1406
Ovr107	1875	GGAAATGGAGGTCATTTGACCTGCCAGGCGCCCTTCGCAAAGAGTGACGA	1924
439_001.nt.6	1407	GGAAATGGAGGTCATTTGACCTGCCAGGCGCCCTTCGCAAAGAGTGACGA	1456

FIGURE 5 (continued)

Ovr107	1925	GGCCCCGTGGGAGAACGGACTCCTCAGACTCTCCCCAATAGCGGAAGTCG	1974
439_001.nt.6	1457	GGCCCCGTGGGAGAACGGACTCCTCAGACTCTCCCCAATAGCGGAAGTCG	1506
Ovr107	1975	ATCTTCTGAAGGATGGCCAATCTGCTCCGGCCCTGGTCTTCCCCCATCCC	2024
439_001.nt.6	1507	ATCTTCTGAAGGATGGCCAATCTGCTCCGGCCCTGGTCTTCCCCCATCCC	1556
Ovr107	2025	GGTGGACAGACTTAACGATCCTTGCTGCAGTCCCTCCGGAGAGGATCTGG	2074
439_001.nt.6	1557	GGTGGACAGACTTAACGATCCTTGCTGCAGTCCCTCCGGAGAGGATCTGG	1606
Ovr107	2075	ACTGGCTGGGAGTGGGGAGGGCGTGGAGACAGTCTACGGAAAGCGCTAGC	2124
439_001.nt.6	1607	ACTGGCTGGGAGTGGGGAGGGCGTGGAGACAGTCTACGGAAAGCGCTAGC	1656
Ovr107	2125	AGACCCCCGAGAGGGTGCAGTGGAGCCCTGAGCATTGTAATATGCGGCCC	2174
439_001.nt.6	1657	AGACCCCCGAGAGGGTGCAGTGGAGCCCTGAGCATTGTAATATGCGGCCC	1706
Ovr107	2175	AGCCTATAAACAGCCTCCGTGCTTAGCAAAAAAAAAAAAAAAAAAAAAA	2221
439_001.nt.6	1707	AGCCTATAAACAGCCTCCGTGCTTAGCAGAAAAAAAAAAAAAAAAAAAAA	1756
Ovr107	2222		2221
439_001.nt.6	1757	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACATACAAAAATAAAAGA	1806
Ovr107	2222		2221
439_001.nt.6	1807	ATAGTCAACAAACAAAATAAGAACTATAGATAATATAAAAATGAAAATA	1856
Ovr107	2222		2221
439_001.nt.6	1857	AAAAAGAGATGGGGTGGGGCCCTTGCTTTACTCTCTCCCTCTGGAGTGG	1906
Ovr107	2222		2221
439_001.nt.6	1907	GCACACTATATTATTTTCGCCCTCCCCCTCTTTTTTTGTATGAGAGGGCTC	1956
Ovr107	2222	2221	
439_001.nt.6	1957	TTTTA	1961

FIGURE 6

Seq. #	Seq. Name	Seq. ID	Seq. Text	Seq. Length
1	Ovr107	1	AGAGCAAGGAAGGGCAGGGGACCTGGGAAGGAAGTTCTGGAAGGCAGTGG	50
2	439_001.nt.8	1		0
3	Ovr107	51	GGTTTGAGATTGGACCCAGGGTCAAGATAGAACATGAAGGTGGGATGAGG	100
4	439_001.nt.8	1		0
5	Ovr107	101	ACATGAACAGAACATGGCCAAGAAGGATCTGGGGGAGCAGCCAGGACGAG	150
6	439_001.nt.8	1		0
7	Ovr107	151	GCGGAGCTGATCCGAGAGGACATCCAGGGGGCTCTGCACAATTACCGCTC	200
8	439_001.nt.8	1		0
9	Ovr107	201	GGGCCGCGGGGAGCGCAGGGCGGCGGCGCTCAGGGCCACGCAGGAGGAGT	250
10	439_001.nt.8	1		0
11	Ovr107	251	TGCAGCGCGACCGCTCGCCCGCCGCTGAGACCCCGCCCCCTGCAGCGCCGC	300
12	439_001.nt.8	1		0
13	Ovr107	301	CCGTCAGTCCGCGCAGTGATCAGCACCGTAGAGCGGGGCGGGGCCGCGG	350
14	439_001.nt.8	1		0
15	Ovr107	351	ACGACCCCAGGCGAAGCCCATTCCTCCGAGGCAGAGGAGGCGCAGAGGCCTG	400
16	439_001.nt.8	1		0
17	Ovr107	401	AGCCGGTGGGGACCTCGAGCAA-----CGCTGACTC----GGC-CTCC	438
18	439_001.nt.8	1	. . GATCTCTTCCAAATGTCCCCGCT--CTCCCCAGGCTCTCC	38
19	Ovr107	439	CCGGACCTGGGTCC-----CCGGGGTCTCTGACCTGGCGGTTCTGCA	479
20	439_001.nt.8	39	. CC-----TCCCGCCACTTGCCAGGG--CTGACCT-----CA	67
21	Ovr107	480	-GGCGGAGCGGGAAAGTGGACATCCTGAACCACGTGTTCGACGACGTAGAG	528
22	439_001.nt.8	68 CCGC-----CAT-CTTAACCGGGTGTCC-----	89
23	Ovr107	529	AGCTTTGTATCGAGGCTGCAGAAGTCGGCGGAGGCGGCCAGGGTGTGGA	578
24	439_001.nt.8	90 ACCTCTCT-----CTGC-----CTGCC-TGGTGTGCTGCC	116
25	Ovr107	579	GCACCGGGAACGCGGCCGCGAGGAGCCGCGCCGGCGCGGCT----GGGGAG	624
26	439_001.nt.8	117 CC-----CGCGTCCCCA-TCGCC-GCGCCCGTCTGCTCCCCCTCAGAG	156

FIGURE 6 (continued)

Ovr107	625	GGCTTGCTGACGCTGCGGGCCAAGCCGCCCTCGGAGGCCGAGTACACCGA	674
439_001.nt.8	157	GGCTTGCTGACGCTGCGGGCCAAGCCGCCCTCGGAGGCCGAGTACACCGA	206
Ovr107	675	CGTGCTGCAGAAGATCAAGTACGCCTTCAGCCTGCTGGCCCGGCTGCGCG	724
439_001.nt.8	207	CGTGCTGCAGAAGATCAAGTACGCCTTCAGCCTGCTGGCCCGGCTGCGCG	256
Ovr107	725	GCAACATCGCCGACCCCTCCTCTCCGGAGCTGTTGCAC TTCCTTTTCGGG	774
439_001.nt.8	257	GCAACATCGCCGACCCCTCCTCTCCGGAGCTGTTGCAC TTCCTTTTCGGG	306
Ovr107	775	CCTCTGCAGATGATTGTGAACACGTCGGGGGGGCCGGAGTTCGCGAGCAG	824
439_001.nt.8	307	CCTCTGCAGATGATTGTGAACACGTCGGGGGGGCCGGAGTTCGCGAGCAG	356
Ovr107	825	TGTGCGGCGGCCGCATCTGACATCGGATGCCGTGGCGCTGCTGCGGGACA	874
439_001.nt.8	357	TGTGCGGCGGCCGCATCTGACATCGGATGCCGTGGCGCTGCTGCGGGACA	406
Ovr107	875	ACGTCACTCCACGTGAAAACGAGCTCTGGACCTCGCTGGGGGACTCGTGG	924
439_001.nt.8	407	ACGTCACTCCACGTGAAAACGAGCTCTGGACCTCGCTGGGGGACTCGTGG	456
Ovr107	925	ACCCGCCCCGGGCTGGAGCTGTCCCCGGAGGAGGGACCCCATACAGACC	974
439_001.nt.8	457	ACCCGCCCCGGGCTGGAGCTGTCCCCGGAGGAGGGACCCCATACAGACC	506
Ovr107	975	CGAGTTCTTCAGCGGCTGGGAGCCGCCGGTCACTGACCCGCAGAGCCGCG	1024
439_001.nt.8	507	CGAGTTCTTCAGCGGCTGGGAGCCGCCGGTCACTGACCCGCAGAGCCGCG	556
Ovr107	1025	CCTGGGAGGACCCAGTTGAGAAACAGCTACAGCACGAGCGGAGGCCCGG	1074
439_001.nt.8	557	CCTGGGAGGACCCAGTTGAGAAACAGCTACAGCACGAGCGGAGGCCCGG	606
Ovr107	1075	CAGCAAAGCGCCCCCAGGTCGCTGTCAATGGTCACCGAGACTTGGAGCC	1124
439_001.nt.8	607	CAGCAAAGCGCCCCCAGGTCGCTGTCAATGGTCACCGAGACTTGGAGCC	656
Ovr107	1125	AGAATCTGAGCCTCAGCTGGAGTCAGAGACAGCAGGAAAATGGGTCCCTGT	1174
439_001.nt.8	657	AGAATCTGAGCCTCAGCTGGAGTCAGAGACAGCAGGAAAATGGGTCCCTGT	706
Ovr107	1175	GTAATTATGACTTCCAGGCCCGCAACAGCAGTGAGCTGTCCGGTCAAGCAG	1224
439_001.nt.8	707	GTAATTATGACTTCCAGGCCCGCAACAGCAGTGAGCTGTCCGGTCAAGCAG	756
Ovr107	1225	CGGGACGTACTGGAGGTCTTGATGACAGTCGTAAGTGGTGGAAAGGTTCCG	1274
439_001.nt.8	757	CGGGACGTACTGGAGGTCTTGATGACAGTCGTAAGTGGTGGAAAGGTTCCG	806

FIGURE 6 (continued)

Ovr107	1275	GGACCCAGCGGGGCGAGGAGGGATATGTGCCCTACAACATCCTGACACCCT	1324
439_001.nt.8	807		856
Ovr107	1325	ACCCCGGACCCCGGGCTGCACCACAGCCAAAGCCCTGCCCGCAGCCTGAAC	1374
439_001.nt.8	857		906
Ovr107	1375	AGCACTCCTCCTCCACCACCAGCCCCAGCCCCGGCCCCACCTCCAGCTCT	1424
439_001.nt.8	907		956
Ovr107	1425	GGCTCGGCCCCGCTGGGACAGGCCCGCTGGGACAGCTGCGATAGCCTCA	1474
439_001.nt.8	957		1006
Ovr107	1475	ACGGCTTGGACCCAGCGAGAAGGAGAAATTCTCCAGATGCTCATCGTC	1524
439_001.nt.8	1007		1056
Ovr107	1525	AACGAGGAAGTGCAGGCGCGCTGGCCCAGGGCCGCTCGGGACCGAGCCG	1574
439_001.nt.8	1057		1106
Ovr107	1575	CGCAGTCCCAGGGCCCCGCGCCCCGGAACCGCAGCTCAGCCCGGGCTCGG	1624
439_001.nt.8	1107		1156
Ovr107	1625	ACGCCTCCGAGGTCCGCGCCTGGCTGCAGGCCAAGGGCTTTAGCTCCGGG	1674
439_001.nt.8	1157		1206
Ovr107	1675	ACCGTGGACGCGCTGGGTGTGCTGACCGGGGCGCAGCTTTTCTCGCTGCA	1724
439_001.nt.8	1207		1256
Ovr107	1725	GAGGGAGGAGCTGCGGGCGGTGAGCCCCGAGGAGGGGGCACGTGTGTACA	1774
439_001.nt.8	1257	.	1306
Ovr107	1775	GCCAGGTACCGTGCAGCGCTCGCTGCTGGAGGACAAAGAGAAAGTGTC	1824
439_001.nt.8	1307		1356
Ovr107	1825	GAGCTGGAGGCAGTGATGGAGAAGCAAAGAAGGTGGAAGGCCGAGGT	1874
439_001.nt.8	1357		1406
Ovr107	1875	GGAAATGGAGGTCATTTGACCTGCCAGGCGCCCTTCGCAAAGAGTGACGA	1924
439_001.nt.8	1407		1456

FIGURE 6 (continued)

Ovr107	1925	GGCCCCGTGGGAGAACGGACTCCTCAGACTCTCCCAATAGCGGAAGTCG	1974
439_001.nt.8	1457		1506
Ovr107	1975	ATCTTCTGAAGGATGGCCAATCTGCTCCGGCCCTGGTCTTCCCCCATCCC	2024
439_001.nt.8	1507		1556
Ovr107	2025	GGTGGACAGACTTAACGATCCTTGCTGCAGTCCCTCCGGAGAGGATCTGG	2074
439_001.nt.8	1557		1606
Ovr107	2075	ACTGGCTGGGAGTGGGGAGGGCGTGGAGACAGTCTACGGAAAGCGCTAGC	2124
439_001.nt.8	1607		1656
Ovr107	2125	AGACCCCCGAGAGGGTGCAGTGGAGCCCTGAGCATTGTAATATGCGGCCC	2174
439_001.nt.8	1657		1706
Ovr107	2175	AGCCTATAAACAGCCTCCGTGCTTAGCAAAAAAAAAAAAAA-----	2215
439_001.nt.8	1707		1756
Ovr107	2216	AAAAAA	2221
439_001.nt.8	1757	.	1806
Ovr107	2222	AACAAACGTTTGGGGTATTCCATGGCCAATACCGTTGTTCCCGTGTGTGA	2221
439_001.nt.8	1807	ACATTGTTATTTTCAGCTCACATTTCCACAGTATTGGAACAACACATCAT	1856
Ovr107	2222		2221
439_001.nt.8	1857	ACCACACACACACAGAACCAATCGAGATATATAAACCCAATGCACACTCA	1906
Ovr107	2222	2221	
439_001.nt.8	1907	AACACCTAAT	1916

FIGURE 7

Ovr107	1	AGAGCAAGGAAGGGCAGGGGACCTGGGAAGGAAGTTCTGGAAGGCAGTGG	50
439_001.nt.10	1		0
Ovr107	51	GGTTTGAGATTGGACCCAGGGTCAAGATAGAACATGAAGGTGGGATGAGG	100
439_001.nt.10	1		0
Ovr107	101	ACATGAACAGAACATGGCCAAGAAGGATCTGGGGGAGCAGCCAGGACGAG	150
439_001.nt.10	1		0
Ovr107	151	GCGGAGCTGATCCGAGAGGACATCCAGGGGGCTCTGCACAATTACCGCTC	200
439_001.nt.10	1		0
Ovr107	201	GGGCCGCGGGGAGCGCAGGGCGGCGGCTCAGGGCCACGCAGGAGGAGT	250
439_001.nt.10	1		0
Ovr107	251	TGCAGCGCGACCGCTCGCCCCGCGCTGAGACCCGCCCCCTGCAGCGCCGC	300
439_001.nt.10	1		0
Ovr107	301	CCGTCAGTCCGCGCAGTGATCAGCACCGTAGAGCGGGGCGGGGCCGCGG	350
439_001.nt.10	1		0
Ovr107	351	ACGACCCCAGGCGAAGCCCATTCCCGAGGCAGAGGAGGCGCAGAGGCCTG	400
439_001.nt.10	1		0
Ovr107	401	AGCCGGTGGGGACCTCGAGCAA-----CGCTGACTC-----GGC-CTCC	438
439_001.nt.10	1 GATCTCTTCCAAATGTCCCCGCT--CTCCCCAGGCTCTCC	38
Ovr107	439	CCGGACCTGGGTCC-----CCGGGGTCCTGACCTGGCGGTTCTGCA	479
439_001.nt.10	39	. CC-----TCCCGCCACTTGCCAGGG--CTGACCT-----CA	67
Ovr107	480	-GGCGGAGCGGGAGTGGACATCCTGAACCACGTGTTGACGACGCTAGAG	528
439_001.nt.10	68 CCGC-----CAT-CTTAACGGGTGTCC-----	89
Ovr107	529	AGCTTTGTATCGAGGCTGCAGAAGTCGGCGGAGGCGGCCAGGGTGCTGGA	578
439_001.nt.10	90 ACCTCTCT-----CTGC-----CTGCC-TGGTGCTGGC	116
Ovr107	579	GCACCGGGAACGCGGCCGCGAGGACCGCGCGGGCGGGCT----GGGGAG	624
439_001.nt.10	117 CC-----CGCGTCCCA-TCGCC-GCGCCGCTCTGCTCCCCCTCAGAG	156

FIGURE 7 (continued)

Ovr107	625	GGCTTGCTGACGCTGCGGGCCAAGCCGCCCTCGGAGGCCGAGTACACCGA	674
439_001.nt.10	157		206
Ovr107	675	CGTGCTGCAGAAGATCAAGTACGCCTTCAGCCTGCTGGCCCGGCTGCGCG	724
439_001.nt.10	207		256
Ovr107	725	GCAACATCGCCGACCCCTCCTCTCCGGAGCTGTTGCACTTCCTTTTCGGG	774
439_001.nt.10	257		306
Ovr107	775	CCTCTGCAGATGATTGTGAACACGTCGGGGGGGCCGGAGTTCGCGAGCAG	824
439_001.nt.10	307		356
Ovr107	825	TGTGCGGCGGCCGCATCTGACATCGGATGCCGTGGCGCTGCTGCGGGACA	874
439_001.nt.10	357		406
Ovr107	875	ACGTCACTCCACGTGAAAACGAGCTCTGGACCTCGCTGGGGGACTCGTGG	924
439_001.nt.10	407		456
Ovr107	925	ACCCGCCCCGGGCTGGAGCTGTCCCCGGAGGAGGGACCCCCATACAGACC	974
439_001.nt.10	457		506
Ovr107	975	CGAGTTCTTCAGCGGCTGGGAGCCGCCGGTCACTGACCCGCAGAGCCGCG	1024
439_001.nt.10	507		556
Ovr107	1025	CCTGGGAGGACCCAGTTGAGAAACAGCTACAGCACGAGCGGAGGCGCCGG	1074
439_001.nt.10	557		606
Ovr107	1075	CAGCAAAGCGCCCCCAGGTCGCTGTCAATGGTCACCGAGACTTGGAGCC	1124
439_001.nt.10	607		656
Ovr107	1125	AGAATCTGAGCCTCAGCTGGAGTCAGAGACAGCAGGAAAATGGGTCTCTGT	1174
439_001.nt.10	657		706
Ovr107	1175	GTAATTATGACTTCCAGGCCCGCAACAGCAGTGAGCTGTCCGGTCAAGCAG	1224
439_001.nt.10	707		756
Ovr107	1225	CGGGACGTACTGGAGGTCCTGGATGACAGTCGTAAGTGGTGGAAGGTTTCG	1274
439_001.nt.10	757		806

FIGURE 7 (continued)

Ovr107	1275	GGACCCAGCGGGGCAGGAGGGATATGTGCCCTACAACATCCTGACACCCCT	1324
439_001.nt.10	807	GGACCCAGCGGGGCAGGAGGGATATGTGCCCTACAACATCCTGACACCCCT	856
Ovr107	1325	ACCCCGGACCCCGGCTGCACCACAGCCAAAGCCCTGCCCGCAGCCTGAAC	1374
439_001.nt.10	857	ACCCCGGACCCCGGCTGCACCACAGCCAAAGCCCTGCCCGCAGCCTGAAC	906
Ovr107	1375	AGCACTCCTCCTCCACCACCAGCCCCAGCCCCGGCCCCACCTCCAGCTCT	1424
439_001.nt.10	907	AGCACTCCTCCTCCACCACCAGCCCCAGCCCCGGCCCCACCTCCAGCTCT	956
Ovr107	1425	GGCTCGGCCCCGCTGGGACAGGCCCGCTGGGACAGCTGCGATAGCCTCA	1474
439_001.nt.10	957	GGCTCGGCCCCGCTGGGACAGGCCCGCTGGGACAGCTGCGATAGCCTCA	1006
Ovr107	1475	ACGGCTTGGACCCCAGCGAGAAGGAGAAATTCTCCCAGATGCTCATCGTC	1524
439_001.nt.10	1007	ACGGCTTGGACCCCAGCGAGAAGGAGAAATTCTCCCAGATGCTCATCGTC	1056
Ovr107	1525	AACGAGGAAGTGCAGGCGCGCCTGGCCCAGGGCCGCTCGGGACCGAGCCG	1574
439_001.nt.10	1057	AACGAGGAAGTGCAGGCGCGCCTGGCCCAGGGCCGCTCGGGACCGAGCCG	1106
Ovr107	1575	CGCAGTCCCAGGGCCCCGCGCCCCGGAACCGCAGCTCAGCCCGGGCTCGG	1624
439_001.nt.10	1107	CGCAGTCCCAGGGCCCCGCGCCCCGGAACCGCAGCTCAGCCCGGGCTCGG	1156
Ovr107	1625	ACGCCTCCGAGGTCCGCGCCTGGCTGCAGGCCAAGGGCTTTAGCTCCGGG	1674
439_001.nt.10	1157	ACGCCTCCGAGGTCCGCGCCTGGCTGCAGGCCAAGGGCTTTAGCTCCGGG	1206
Ovr107	1675	ACCGTGGACGCGCTGGGTGTGCTGACCGGGGCGCAGCTTTTCTCGCTGCA	1724
439_001.nt.10	1207	ACCGTGGACGCGCTGGGTGTGCTGACCGGGGCGCAGCTTTTCTCGCTGCA	1256
Ovr107	1725	GAGGGAGGAGCTGCGGGCGGTGAGCCCCGAGGAGGGGGCACGTGTGTACA	1774
		.	
439_001.nt.10	1257	GAAGGAGGAGCTGCGGGCGGTGAGCCCCGAGGAGGGGGCACGTGTGTACA	1306
Ovr107	1775	GCCAGGTACCGTGCAGCGCTCGCTGCTGGAGGACAAAGAGAAAGTGTCA	1824
439_001.nt.10	1307	GCCAGGTACCGTGCAGCGCTCGCTGCTGGAGGACAAAGAGAAAGTGTCA	1356
Ovr107	1825	GAGCTGGAGGCAGTGATGGAGAAGCAAAAGAAGAAGGTGGAAGGCCAGGT	1874
439_001.nt.10	1357	GAGCTGGAGGCAGTGATGGAGAAGCAAAAGAAGAAGGTGGAAGGCCAGGT	1406
Ovr107	1875	GGAAATGGAGGTCAATTTGACCTGCCAGGCGCCCTTCGCAAAGAGTGACGA	1924
439_001.nt.10	1407	GGAAATGGAGGTCAATTTGACCTGCCAGGCGCCCTTCGCAAAGAGTGACGA	1456

FIGURE 7 (continued)

Ovr107	1925	GGCCCCGTGGGAGAACGGACTCCTCAGACTCTCCCCAATAGCGGAAGTCG	1974
439_001.nt.10	1457	GGCCCCGTGGGAGAACGGACTCCTCAGACTCTCCCCAATAGCGGAAGTCG	1506
Ovr107	1975	ATCTTCTGAAGGATGGCCAATCTGCTCCGGCCCTGGTCTTCCCCCATCCC	2024
439_001.nt.10	1507	ATCTTCTGAAGGATGGCCAATCTGCTCCGGCCCTGGTCTTCCCCCATCCC	1556
Ovr107	2025	GGTGGACAGACTTAACGATCCTTGCTGCAGTCCCTCCGGAGAGGATCTGG	2074
439_001.nt.10	1557	GGTGGACAGACTTAACGATCCTTGCTGCAGTCCCTCCGGAGAGGATCTGG	1606
Ovr107	2075	ACTGGCTGGGAGTGGGGAGGGCGTGGAGACAGTCTACGGAAAGCGCTAGC	2124
439_001.nt.10	1607	ACTGGCTGGGAGTGGGGAGGGCGTGGAGACAGTCTACGGAAAGCGCTAGC	1656
Ovr107	2125	AGACCCCCGAGAGGGTGCAGTGGAGCCCTGAGCATTGTAATATGCGGCC	2174
439_001.nt.10	1657	AGACCCCCGAGAGGGTGCAGTGGAGCCCTGAGCATTGTAATATGCGGCC	1706
Ovr107	2175	AGCCTATAAACAGCCTCCGTGCTTAGCAAAAAAAAAAAAAAAAAAAAAA	2221
439_001.nt.10	1707	AGCCTATAAACAGCCTCCGTGCTTAGCAG	1735

FIGURE 8

Ovr110	1	GAGTCACCAAGGAAGGCAGCGGCAGCTCCACTCAGCCAGTACCCAGA	47
Ovr110v1	1	TGTGAGTCACCAAGGAAGGCAGCGGCA-CTCCACTCAGCCAGTACCCAGA	49
Ovr110	48	TACGCTGGGAACCTTCCCCAGCCATGGCTTCCCTGGGGCAGATCCTCTTC	97
Ovr110v1	50	TACGCTGGGAACCTTCCCCAGCCATGGCTTCCCTGGGGCAGATCCTCTTC	99
Ovr110	98	TGGAGCATAATTAGCATCATCATTATTCTGGCTGGAGCAATTGCACTCAT	147
Ovr110v1	100	TGGAGCATAATTAGCATCATCATTATTCTGGCTGGAGCAATTGCACTCAT	149
Ovr110	148	CATTGGCTTTGGTATTT-----	164
Ovr110v1	150	CATTGGCTTTGGTATTTTCAGAAGTCTCTGTCTGGCTTTCAGCAATGAAGG	199
Ovr110	165	-----	164
Ovr110v1	200	GTTTGGTTGTAGAAGTTCCAAGGCTTCCCTTAGCATTGATCTTTGCTTCC	249
Ovr110	165	-----CAGGGAGACACTCCATCACAGTCACTACTGTGCGCTCAGCTGG	207
Ovr110v1	250	TGAACTGCAGGGAGACACTCCATCACAGTCACTACTGTGCGCTCAGCTGG	299
Ovr110	208	GAACATTGGGGAGGATGGAATCCAGAGCTGCACTTTTGAACCTGACATCA	257
Ovr110v1	300	GAACATTGGGGAGGATGGAATCCTGAGCTGCACTTTTGAACCTGACATCA	349
Ovr110	258	AACTTTCTGATATCGTGATACAATGGCTGAAGGAAGGTGTTTTAGGCTTG	307
Ovr110v1	350	AACTTTCTGATATCGTGATACAATGGCTGAAGGAAGGTGTTTTAGGCTTG	399
Ovr110	308	GTCCATGAGTTCAAAGAAGGCAAAGATGAGCTGTCGGAGCAGGATGAAAT	357
Ovr110v1	400	GTCCATGAGTTCAAAGAAGGCAAAGATGAGCTGTCGGAGCAGGATGAAAT	449
Ovr110	358	GTTTCAAGAGGCCGGACAGCAGTGTGCTGATCAAGTGATAGTTGGCAATG	407
Ovr110v1	450	GTTTCAAGAGGCCGGACAGCAGTGTGCTGATCAAGTGATAGTTGGCAATG	499
Ovr110	408	CCTCTTTGCGGCTGAAAAACGTGCAACTCACAGATGCTGGCACCTACAAA	457
Ovr110v1	500	CCTCTTTGCGGCTGAAAAACGTGCAACTCACAGATGCTGGCACCTACAAA	549
Ovr110	458	TGTTATATCATCACTTCTAAAGGCAAGGGGAATGCTAACCTTGAGTATAA	507
Ovr110v1	550	TGTTATATCATCACTTCTAAAGGCAAGGGGAATGCTAACCTTGAGTATAA	599
Ovr110	508	AACTGGAGCCTTCAGCATGCCGGAAGTGAATGTGGACTATAATGCCAGCT	557
Ovr110v1	600	AACTGGAGCCTTCAGCATGCCGGAAGTGAATGTGGACTATAATGCCAGCT	649

FIGURE 8 (continued)

Ovr110	558	CAGAGACCTTGCGGTGTGAGGCTCCCGATGGTTCCCCCAGCCCACAGTG	607
Ovr110v1	650	CAGAGACCTTGCGGTGTGAGGCTCCCGATGGTTCCCCCAGCCCACAGTG	699
Ovr110	608	GTCTGGGCATCCCAAGTTGACCAGGGAGCCAACTTCTCGGAAGTCTCCAA	657
Ovr110v1	700	GTCTGGGCATCCCAAGTTGACCAGGGAGCCAACTTCTCGGAAGTCTCCAA	749
Ovr110	658	TACCAGCTTTGAGCTGAACTCTGAGAATGTGACCATGAAGGTTGTGTCTG	707
Ovr110v1	750	TACCAGCTTTGAGCTGAACTCTGAGAATGTGACCATGAAGGTTGTGTCTG	799
Ovr110	708	TGCTCTACAATGTTACGATCAACAACACATACTCCTGTATGATTGAAAAT	757
Ovr110v1	800	TGCTCTACAATGTTACGATCAACAACACATACTCCTGTATGATTGAAAAT	849
Ovr110	758	GACATTGCCAAAGCAACAGGGGATATCAAAGTGACAGAATCGGAGATCAA	807
Ovr110v1	850	GACATTGCCAAAGCAACAGGGGATATCAAAGTGACAGAATCGGAGATCAA	899
Ovr110	808	AAGGCGGAGTCACCTACAGCTGCTAAACTCAAAGGCTTCTCTGTGTGTCT	857
Ovr110v1	900	AAGGCGGAGTCACCTACAGCTGCTAAACTCAAAGGCTTCTCTGTGTGTCT	949
Ovr110	858	CTTCTTTCTTTGCCATCAGCTGGGCACTTCTGCCTCTCAGCCCTTACCTG	907
Ovr110v1	950	CTTCTTTCTTTGCCATCAGCTGGGCACTTCTGCCTCTCAGCCCTTACCTG	999
Ovr110	908	ATGCTAAAATAATGTGCCTCGGCCACAAAAAGCATGCAAAGTCATTGTT	957
Ovr110v1	1000	ATGCTAAAATAATGTGCCTTGGCCACAAAAAGCATGCAAAGTCATTGTT	1049
Ovr110	958	ACAACAGGGATCTACAGAACTATTTCAACCACCAGATATGACCTAGTTTTA	1007
Ovr110v1	1050	ACAACAGGGATCTACAGAACTATTTCAACCACCAGATATGACCTAGTTTTA	1099
Ovr110	1008	TATTTCTGGGAGGAAATGAATTCATATCTAGAAGTCTGGAGTGAGCAAAC	1057
Ovr110v1	1100	TATTTCTGGGAGGAAATGAATTCATATCTAGAAGTCTGGAGTGAGCAAAC	1149
Ovr110	1058	AAGAGCAAGAAACAAAAAGAAGCCAAAAGCAGAAGGCTCCAATATGAACA	1107
Ovr110v1	1150	AAGAGCAAGAAACAAAAAGAAGCCAAAAGCAGAAGGCTCCAATATGAACA	1199
Ovr110	1108	AGATAAATCTATCTTCAAAGACATATTAGAAGTTGGGAAAATAATTCATG	1157
Ovr110v1	1200	AGATAAATCTATCTTCAAAGACATATTAGAAGTTGGGAAAATAATTCATG	1249
Ovr110	1158	TGAACTAGA-----	1166
Ovr110v1	1250	TGAACTAGACAAGTGTGTTAAGAGTGATAAGTAAAATGCACGTGGAGACA	1299

FIGURE 8 (continued)

Ovr110	1167 -----	1166
Ovr110v1	1300 AGTGCATCCCCAGATCTCAGGGACCTCCCCCTGCCTGTCACCTGGGGAGT	1349
Ovr110	1167 -----	1166
Ovr110v1	1350 GAGAGGACAGGATAGTGCATGTTCTTTGTCTCTGAATTTTATAGTTATATG	1399
Ovr110	1167 -----	1166
Ovr110v1	1400 TGCTGTAATGTTGCTCTGAGGAAGCCCCTGGAAAGTCTATCCCAACATAT	1449
Ovr110	1167 -----	1166
Ovr110v1	1450 CCACATCTTATATTCCACAAATTAAGCTGTAGTATGTACCCTAAGACGCT	1499
Ovr110	1167 -----	1166
Ovr110v1	1500 GCTAATTGACTGCCACTTCGCAACTCAGGGGCGGCTGCATTTTAGTAATG	1549
Ovr110	1167 -----	1166
Ovr110v1	1550 GGTCAAATGATTCACTTTTTATGATGCTTCCAAAGGTGCCTTGGCTTCTC	1599
Ovr110	1167 -----	1166
Ovr110v1	1600 TTCCCAACTGACAAATGCCAAAGTTGAGAAAAATGATCATAATTTTAGCA	1649
Ovr110	1167 -----	1166
Ovr110v1	1650 TAAACAGAGCAGTCGGCGACACCGATTTTATAAATAAACTGAGCACCTTC	1699
Ovr110	1167 -----	1166
Ovr110v1	1700 TTTTAAACAAACAAATGCGGGTTTATTTCTCAGATGATGTTTCATCCGTG	1749
Ovr110	1167 -----	1166
Ovr110v1	1750 AATGGTCCAGGGAAGGACCTTTCACCTTGACTATATGGCATTATGTCATC	1799
Ovr110	1167 -----	1166
Ovr110v1	1800 ACAAGCTCTGAGGCTTCTCCTTTCCATCCTGCGTGGACAGCTAAGACCTC	1849
Ovr110	1167 -----	1166
Ovr110v1	1850 AGTTTTCAATAGCATCTAGAGCAGTGGGACTCAGCTGGGGTGATTTGCGC	1899
Ovr110	1167 -----	1166
Ovr110v1	1900 CCCCATCTCCGGGGGAATGCTGAAGACAATTTGGTTACCTCAATGAGG	1949

FIGURE 8 (continued)

Ovr110	1167	-----	1166
Ovr110v1	1950	GAGTGGAGGAGGATACAGTGCTACTACCAACTAGTGGATAAAGGCCAGGG	1999
Ovr110	1167	-----	1166
Ovr110v1	2000	ATGCTGCTCAACCTCCTACCATGTACAGGACGTCTCCCCATTACAACCTAC	2049
Ovr110	1167	-----GTCAACTGTGTTCAGGGCTAAGAAACCCTGGTTTTGAGT	1204
Ovr110v1	2050	CCAATCCGAAGTGTCAACTGTGTTCAGGACTAAGAAACCCTGGTTTTGAGT	2099
Ovr110	1205	AGAAAAGGGCCTGGAAAGAGGGGAGCCAAACAATCTGTCTGCTTCCTCAC	1254
Ovr110v1	2100	AGAAAAGGGCCTGGAAAGAGGGGAGCCAAACAATCTGTCTGCTTCCTCAC	2149
Ovr110	1255	ATTAGTCATTGGCAAATAAGCATTCTGTCTCTTTGGCTGCTGCCTCAGCA	1304
Ovr110v1	2150	ATTAGTCATTGGCAAATAAGCATTCTGTCTCTTTGGCTGCTGCCTCAGCA	2199
Ovr110	1305	CAGAGAGCCAGAACTCTATCGGGCACCAGGATAACATCTCTCAGTGAACA	1354
Ovr110v1	2200	CAGAGAGCCAGAACTCTATCGGGCACCAGGATAACATCTCTCAGTGAACA	2249
Ovr110	1355	GAGTTGACAAGGCCTATGGGAAATGCCTGATGGGATTATCTTCAGCTTGT	1404
Ovr110v1	2250	GAGTTGACAAGGCCTATGGGAAATGCCTGATGGGATTATCTTCAGCTTGT	2299
Ovr110	1405	TGAGCTTCTAAGTTTCTTTCCCTTCATTCTACCCCTGCAAGCCAAGTTCTG	1454
Ovr110v1	2300	TGAGCTTCTAAGTTTCTTTCCCTTCATTCTACCCCTGCAAGCCAAGTTCTG	2349
Ovr110	1455	TAAGAGAAATGCCTGAGTTCAGCTCAGGTTTTCTTACTCTGAATTTAGA	1504
Ovr110v1	2350	TAAGAGAAATGCCTGAGTTCAGCTCAGGTTTTCTTACTCTGAATTTAGA	2399
Ovr110	1505	TCTCCAGACCCTGCCTGGCCACAATTCAAATTAAGGCAACAAACATATAC	1554
Ovr110v1	2400	TCTCCAGACCCTTCCTGGCCACAATTCAAATTAAGGCAACAAACATATAC	2449
Ovr110	1555	CTTCCATGAAGCACACACAGACTTTTGAAAGCAAGGACAATGACTGCTTG	1604
Ovr110v1	2450	CTTCCATGAAGCACACACAGACTTTTGAAAGCAAGGACAATGACTGCTTG	2499
Ovr110	1605	AATTGAGGCCTTGAGGAATGAAGCTTTGAAGGAAAAGAATACTTTGTTTC	1654
Ovr110v1	2500	AATTGAGGCCTTGAGGAATGAAGCTTTGAAGGAAAAGAATACTTTGTTTC	2549
Ovr110	1655	CAGCCCCCTTCCACACTCTTCATGTGTTAACCCTGCCTTCCTGGACCT	1704
Ovr110v1	2550	CAGCCCCCTTCCACACTCTTCATGTGTTAACCCTGCCTTCCTGGACCT	2599

FIGURE 8 (continued)

Ovr110	1705	TGGAGCCACGGTGACTGTATTACATGTTGTTATAGAAAACTGATTTTGA	1754
Ovr110v1	2600	TGGAGCCACGGTGACTGTATTACATGTTGTTATAGAAAACTGATTTTGA	2649
Ovr110	1755	GTTCTGATCGTTCAAGAGAATGATTAAATATACATTTCCTAAAAAAAAA	1804
Ovr110v1	2650	GTTCTGATCGTTCAAGAGAATGATTAAATATACATTTCCTA	2690
Ovr110	1805	AAAAAA	1811
Ovr110v1	2691		2690

FIGURE 9

Ovr110	1 MASLGQILFWSIISIIIIILAGAIALIIGFGISGRHSITVTTVASAGNIGE	50
	
439_002.aa.2	1 MASLGQILFWSIISIIIIILAGAIALIIGFGIS---EVS VWLSAMKGLVVE	47
Ovr110	51 DGIQSCTFEPDIKLSDIVIQWLKEGVLGLVHEFKEGKDELSEQDEMFRGR	100
	
439_002.aa.2	48 -----VPR LPLALIFAS	59
Ovr110	101 TAVFADQVIVGNASLR LKNVQLTDAGTYKCYIITSRGKGNANLEYKTGAF	150
439_002.aa.2	60	59
Ovr110	151 SMPEVNVDYNASSETLRCEAPRWFPQPTVVWASQVDQGANFSEVSNTSFE	200
439_002.aa.2	60	59
Ovr110	201 LNSENVTMKVSVSVLYNVTINNTYSCMIENDIAKATGDIKVTSEIKRRSH	250
439_002.aa.2	60	59
Ovr110	251 LQLLSKASLCVSSFFAISWALLPLSPYIMLK	282
439_002.aa.2	60	59

FIGURE 10

Ovr110	1	MASLGQILFWSIISIIIIILAGAIALIIGFGISGRHSITVTTVASAGNIGE	50
439_002.aa.3	1	..:	20
Ovr110	51	DGIQSCTFEFDIKLSDIVIQWLKEGVLGLVHEFKEGKDELSEQDEMFRGR	100
439_002.aa.3	21	.	70
Ovr110	101	DGILSCTFEFDIKLSDIVIQWLKEGVLGLVHEFKEGKDELSEQDEMFRGR	150
439_002.aa.3	71		120
Ovr110	151	TAVFADQVIVGNASRLRKNVQLTDAGTYKCYIITSKGRGNANLEYKTGAF	200
439_002.aa.3	71		120
Ovr110	151	TAVFADQVIVGNASRLRKNVQLTDAGTYKCYIITSKGRGNANLEYKTGAF	200
439_002.aa.3	121		170
Ovr110	201	SMPEVNVDYNASSETLRCEAPRWFPQPTVVWASQVDQGANFSEVSNTSFE	250
439_002.aa.3	171		220
Ovr110	201	SMPEVNVDYNASSETLRCEAPRWFPQPTVVWASQVDQGANFSEVSNTSFE	250
439_002.aa.3	171		220
Ovr110	251	LNSENVTMKVSVLYNVTINNTYSCMIENDIAKATGDIKVTSEIKRRSH	282
439_002.aa.3	221		252
Ovr110	251	LNSENVTMKVSVLYNVTINNTYSCMIENDIAKATGDIKVTSEIKRRSH	282
439_002.aa.3	221		252

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.